

Thank you for the reviewers' comments and the opportunity for us to improve this paper. In the past few weeks, we have revised this manuscript and highlighted the relevant changes. Revisions have been made to incorporate all of the reviewers' concerns with point-by-point responses. The new version has implemented as many improvements as possible.

Response to Comments from Reviewer 2

The authors proposed proposes a deep ensemble learning to improve performance. This work is meaningful and some comments below may help authors to improve the presentation of this article.

General Response

The authors would like to thank the reviewer for the professional comments and careful examination. The article is revised and reorganized based on the constructive comments and suggestions given by the reviewers. Please see the point-by-point responses for details.

Comment 2-1

The Introduction part can be simple and core clearly, you should write concise and explicit about your main contributions.

Response

We have carefully reviewed the introduction. The introduction was condensed to keep the relevant information. The introduction part has also been modified to incorporate the newest COVID-CT approaches and references as per comments 2-12.

Comment 2-2

The proposed method in this paper is only used in the COVID dataset and has achieved good results. Please explain the generalization and robust performance of the proposed framework.

Response

Ensemble modelling of Deep CNN models has been proved to improve the classification performance, when compared to a single model, by lowering the generalization error as the ensemble can meta-learn from a broader set of independent features [8]. The purpose of employing Ensemble models is to reduce the generalization error on the test database by reducing the variance. Ensemble modelling had been used in various fields and found to produce enhanced results [9]. The presented study only extends this modelling and is limited to Ensemble learning strategies on Deep CNNs for image classification. Hence image classification capability of the proposed model can be further explored in different applications like forest fire detection, disease identification, cancer prediction etc. Authors believe that the proposed model may get robust performance and better generalization compared to single CNN models in such applications. (mentioned in different parts of the revised version and can notice page 17 including section 3.9)

Comment 2-3

The authors propose a multiple method ensemble, which is derived from the fusion of multiple algorithms. Please explain the theoretical computational complexity of the framework.

Response

In Shallow and Deep Meta-Learning Ensemble strategy, the feature maps obtained from the final convolutional layer of the four trained Deep CNN models were used for meta-learning. For training the meta learning models these feature maps are used as input rather than the CT scan images. The feature maps are available after training the Deep CNNs. The time complexity of the framework includes time for training four base CNNs as well as training the meta-learners. The space complexity of the frameworks includes storing the CT scan images as input for Deep CNNs and also the output deep feature maps which will then serve as input for the meta learners. The trained parameters of the base learners and the meta learners need to be saved for inference. Table 6 (in the revised manuscript) shows the total parameters of the presented models. During inference only these parameters are required in the memory.

The presented Deep and Shallow Meta-Learning Ensemble CNNs models employed meta-learners after the convolutional layers and hence contained lesser number of parameters. The Attentive Ensemble took the longest time for inference due to the computationally heavy attention mechanism

The above information is included in the revised manuscript (Page Nos: 20, 21, 22 and 23)

Comment 2-4

The authors use CNN for global feature extraction and texture feature extraction, more details can be added.

Response

Add to section 2.2 page 6

Convolutional Neural Networks (CNNs) have been traditionally used for extracting deep features from image datasets [47] and have been extended to other machine learning tasks due to its ability to capture complex and positionally invariant feature maps. CNNs train convolutional kernels to extract simple low-level features in shallow layers and complex high-level features in deeper layers. CNNs make use of local receptive fields, in which hidden units are connected to local patches of the lower layer. This allows the network to capture local spatial relationships of the image pixels. Weight sharing of the kernels enables translation-invariance of the network to objects in images. It also reduces the number of parameters of the network. CNNs also use pooling as a parameter free down sampling operation serving to aggregate information. Section 2.2 discusses different CNN architectures used in this work such as VGG-19, Inception-V3, ResNet-152-v2, InceptionResNet-V2, DenseNet-201, XceptionNet and EfficientNet-B7.

Comment 2-5

What is the initialized parameters for meta-learning, and how to choose best?

Response

The Ensemble strategies presented in this study use the output generated by four trained base CNNs for meta learning. In the Voting meta-learning, the softmax scores obtained from the four trained classification models corresponding to each class were maxed or averaged to obtain the Ensemble model's predicted class. In the Shallow and Deep Meta-Learning Ensemble strategy, the feature maps obtained from the final convolutional layer of the four trained Deep CNN models were used for meta-learning. The initial hyperparameters for the models are presented in Section 2.4. In Shallow and Deep Meta-Learning Ensemble strategy, the meta-learning models are trained on the CNN feature outputs. The parameters of the model are optimized while training. Therefore, the initial parameters of the models are randomly set.

Comment 2-6

Some core equations of proposed method are necessary. The optimize flow may display by pseudo-code.

Response

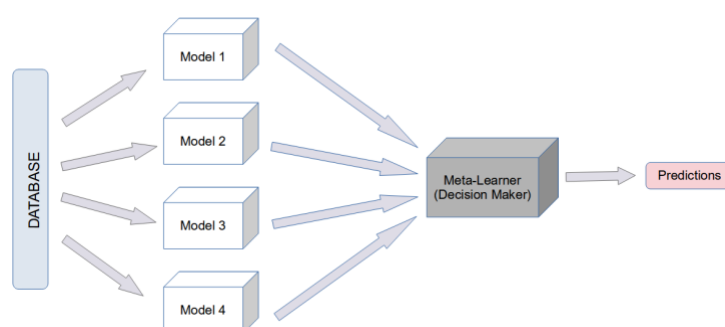
The comment has been addressed in Section 2.3 of the revised manuscript (Page No. 7). Since we are proposing 3 different sets of models (each model contains different algorithms), it is very difficult to insert Pseudo code for all methods because of the space constraints. We have incorporated details of the proposed method with more clarity.

Comment 2-7

Why choose meta optimization, not including grid search, random search, genetic algorithm, particle swarm optimization. How to ensure optimization is robustness.

Response

Ensemble modelling is an approach where multiple separate base models are developed to predict an outcome. The meta-learners then aggregate each base model's predictions to provide a single final prediction. The ensemble modelling of Deep CNN models has been proved to improve the classification performance, when compared to a single model, by lowering the generalization error as the ensemble can meta-learn from a broader set of independent features [8]. Ensemble models are used to reduce the generalization error on the test database by reducing the variance.



Ensemble Learning Model

In this study, meta-learners are used to combine the four Deep CNN outputs. The Deep CNN models were trained using Adam optimizer for convergence. These meta-learners are optimized by training them on the output of the Deep CNN. The methods in the review comment (grid search, random search, genetic algorithm, particles swarm optimization) are search strategies or heuristic strategies used for hyper-parameter optimization and not for synergizing in Ensemble learning (modified in page 3).

Comment 2-8

In the experimental part of section 4, what is way to select a part of dataset for meta learning.

Response

95% of the database was used for 5-fold cross validation. The rest 5% of the database was used for testing. In the 5-fold cross validation dataset, 80% was used as training and 20% was used for cross validation during each fold. Table 1 (page 13) shows the database split. A pseudo random split generator was used to obtain the train, validation and test datasets. This allows the proposed model to trained and evaluated on the same data splits.

Comment 2-9

The proposed method only runs a dataset, the performances are better other approaches, however, how to verify the proposed method is robustness? This web gives 6752 samples and 617,775 slices. Chest CT images are deposited into the China National Center for Bioinformation at the website (<http://ncov-ai.big.ac.cn/download?lang=en>).

Response

The robustness of the proposed framework to generalization was tested by evaluating the trained model on a subset of a Chinese CT Scan database [11]. 1470 CT scan images were tested to classify them as COVID infection, common pneumonia or Normal condition. Table 6 shows the accuracy and F1 score of the evaluated models on Chinese CT Scan database. The Deep CNN models struggled to classify some images due to the variance in the Chinese data distribution compared to training database. The proposed models could perform decently, ensuring the Ensemble approach's generalisation capability.

Table Evaluation Metrics on Chinese CT Scan database

Model	Accuracy	F1 Score
VGG-19	0.8880	0.8879
Inception-v3	0.8977	0.8974
ResNet-152-v2	0.9253	0.9251
InceptionResNet-v2	0.9499	0.9498
DenseNet-201	0.9482	0.9481
Xception	0.9511	0.9508
EfficientNet-b7	0.8794	0.8794
Average Ensemble CNNs	0.9641	0.9641
Majority Ensemble CNNs	0.9710	0.9698
Logistic Ensemble CNNs	0.9402	0.9401
SVC Ensemble CNNs	0.9395	0.9394
Decision Tree Ensemble CNNs	0.9513	0.9511
Naïve Bayes Ensemble CNNs	0.9381	0.9369
KNN Ensemble CNNs	0.9290	0.9287
Random Forest Ensemble CNNs	0.9644	0.9643
Bagging Ensemble CNNs	0.9597	0.9597
AdaBoost Ensemble CNNs	0.9644	0.9643
Gradient Boosting Ensemble CNNs	0.9750	0.9747
NN Ensemble CNNs	0.9766	0.9765
FCN Ensemble CNNs	0.9749	0.9749
Attentive Ensemble CNNs	0.9784	0.9782

Comment 2-10

In Section 3, the author only verified the value of the best performance obtained in the experiment {27,...,34}, please conduct comparative experiments with more values.

Response

The mentioned references propose different Deep CNN architectures namely VGG [1], Inception [2], ResNet [3], InceptionResNet [4], DenseNet [5], XceptionNet [6] and EfficientNet [7]. Various pretrained versions of each of these architectures are made publicly available by the corresponding authors. VGG-19, Inception-V3, ResNet-152-v2, InceptionResNet-V2, DenseNet-201, XceptionNet and EfficientNet-B7 were used in this study as they were presented as the latest versions encompassing the state-of-the-art CNN design. This study aims to select only four base models for Ensemble learning. Comparative experiments of seven Deep CNN models are already presented in this study.

Comment 2-11

The gender or age is a key factor, considering it may have new gains.

Response

The presented study is limited to Ensemble learning strategies on Deep CNNs for image classification. The gender and age of the patients are not disclosed in the database used for this study.

Comment 2-12

The newest COVID-CT approaches may add this work as compared learning methods. For example, references' methods can be included.

Response

Latest approaches for COVID-CT scan classification works are incorporated in Section 1 of the revised manuscript.

Comment 2-13

G-mean can be measured imbalance dataset, which is a metric in this work.

Response

F1-Score is the metric used in the work to measure the efficiency of the proposed model when evaluated in the imbalanced dataset. The F1-score calculates the harmonic mean of precision and recall rather than the geometric mean calculated by the G-mean.

Comment 2-14

Descriptions of COVID19 biology significance combining deep learning methods are given in the section of Results or Discussion.

Response

The description of covid 19 biological significance combining deep learning method is added in introduction session.

References

- [1] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." 2015.
- [2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. abs/1512.0, 2015, [Online]. Available: <http://arxiv.org/abs/1512.00567>.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *CoRR*, vol. abs/1512.0, 2015, [Online]. Available: <http://arxiv.org/abs/1512.03385>.
- [4] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *AAAI Conf. Artif. Intell.*, vol. abs/1602.0, 2016, [Online]. Available: <http://arxiv.org/abs/1602.07261>.
- [5] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *CoRR*, vol. abs/1608.0, 2016, [Online]. Available: <http://arxiv.org/abs/1608.06993>.
- [6] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *IEEE Conf. Comput. Vis. Pattern*

- Recognit.*, 2017. 152
- [7] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Int. Conf. Mach. Learn.*, 2019. 153
- [8] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, 1990, doi: 10.1109/34.58871. 154
- [9] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151. 155
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 25, [Online]. Available: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. 156
- [11] K. Zhang *et al.*, "Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography.," *Cell*, vol. 181, no. 6, pp. 1423–1433.e11, Jun. 2020, doi: 10.1016/j.cell.2020.04.045. 157

152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193

The revised version with response to the reviewers.

Deep Feature Meta-Learners Ensemble Models for Covid CT Scan Classification

Jibin B Thomas¹, Shihabudheen KV², Sheik Mohammed Sulthan³ and Adel Al-Jumaily^{4,5,6,7,*}

¹ Department of Electrical Engineering, National Institute of Technology, Calicut, India; jibinbthomas2000@gmail.com

² Department of Electrical Engineering, National Institute of Technology, Calicut, India; shihabudheen@nitc.ac.in

³ Faculty of Engineering, Universiti Teknologi Brunei, Brunei Darussalam; sheik.sulthan@utb.edu.bn

⁴ Faculty of Engineering, Universiti Teknologi Brunei, Brunei Darussalam; adel.al-jumaily@ieee.org

⁵ ENSTA Bretagne, French State Graduate, 29200 Brest, France.

⁶ School of Computing Mathematics and Engineering, Charles Sturt University, Wagga Wagga, NSW, Australia

⁷ School of Science, Edith Cowan University, Joondalup, WA, Australia; adel.al-jumaily@ieee.org

* Correspondence: adel.al-jumaily@ieee.org;

Abstract: The infectious nature of the COVID-19 virus demands rapid detection to quarantine the infected to isolate the spread or provide the necessary treatment if required. Analysis of COVID-19 infected chest Computed Tomography Scans (CT scans) have been shown to be successful in detecting the disease, making it essential in radiology assessment and screening of infected patients. A single model Deep CNN models have been used to extract complex information pertaining to the CT scan images allowing for in-depth analysis and thereby aiding in the diagnosis of the infection by automatically classifying the chest CT scan images as infected or non-infected. The feature maps obtained from the final convolution layer of the Deep CNN models contain complex and positional encoding of the images' features. The ensemble modelling of these Deep CNN models has been proved to improve the classification performance, when compared to a single model, by lowering the generalization error as the ensemble can meta-learn from a broader set of independent features. This paper presents Deep Ensemble Learning models to synergize Deep CNN models by combining these feature maps to create deep feature vectors or deep feature maps that are then trained on meta shallow and deep learners to improve the classification. This paper also proposes a novel Attentive Ensemble Model that utilizes an attention mechanism to focus on significant feature embeddings while learning the Ensemble feature vector. The proposed Attentive Ensemble model provided better generalization, outperforming Deep CNN models and conventional Ensemble learning techniques as well as Shallow and Deep meta-learning Ensemble CNNs models. Radiologists can use the presented automatic Ensemble classification models to assist identify infected chest CT scans and save lives.

Keywords: Deep learning; Covid-19 prediction; Ensemble classification; CT scan images

Citation: To be added by editorial staff during production.

Academic Editor: Firstname
Lastname

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Coronavirus disease (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The surge of the disease since late 2019 evolved into a pandemic affecting billions of people and thousands of sectors. The reverse transcriptase-polymerase chain reaction (RT-PCR) test is regarded as the gold standard for confirming COVID-19 infection [1]. The shortage of RT-PCR kits and

the need for intimate interaction with the patients while testing can pose a major problem for rapid detection and infection discovery. Radiological analysis of Computed Tomography (CT) scans and X-ray chest images have proved effective in finding symptoms caused by the COVID-19 disease in infected patients [2] and can be formed as a first line of instantaneous detection before RT-PCR testing. These images analysis and finding accurate diagnosing required automatic tools with minimum human interruption like machine learning and deep learning. Deep Learning have been gaining popularity in recent years for analysis of unstructured data like medical images with Deep convolutional neural networks (CNNs) being used extensively [3] due to its ability to capture complex and positional features thereby aiding doctors while diagnosing patients.

Many of these state-of-the-art (SOTA) Deep CNN architectures are developed and open sourced by universities, companies, and research institutions for further improvements thus, contributing to the extraordinary pace of machine learning research and development. For domain-specific problems, these models may be retrained or fine-tuned with pre-trained weights available for popular or domain-specific databases. Various deep-learning models have been suggested and evaluated for COVID-19 prediction from CT scans images. CNN based long short-term memory network is used for COVID-19 classification from pneumonia [4]. Better accuracy was obtained in the proposed model compared to SOTA machine learning models. A Genetic Deep Learning Convolutional Neural Network (GDCNN) was applied for effective COVID-19 prediction [5]. The estimation of COVID-19 diagnosis is performed using 3D CT images, which use different levels of ResNet architecture [6]. GoogleNet Inception v3 with transfer learning capability was proposed for efficient covid prediction in [7]. Severity prediction for COVID-19 has been implemented using image segmentation and VB-Net [8]. Several efficient individual models are reported on CT scans images for COVID-19 detection [9-12]. It has been shown that the detection of the most effective model from a group of CNN methods is ambiguous because the task involves several performance metrics. Various proven Deep CNNs like AlexNet, VGG, DenseNet, SqueezeNet, GoogleNet (Inception Net), MobileNet, ResNet, Xception experimented for COVID radiology studies [13-15]. Efficient Nets were used as base feature extractors in [16-18]. These models proved to be useful with transfer learning as they were pre trained on the ImageNet database. Various improvements to the proven Deep CNNs were presented in [19-22] proposed spiking neural networks which gave a F1 score of 0.74 for on CT scan database. These studies utilize a single CNN classifier which can overfit the test database resulting in high variance and low generalization on real-world systems.

It is proved that ensemble learning of different CNN architectures in a suitable format will improve the feature extraction capability of machine learning [23]. Studies have shown that Ensemble modelling of the SOTA Deep CNN architectures produces better generalization on medical image databases compared to a single Deep CNN model [24-27] where Conventional Ensemble techniques in machine learning and Deep CNN make use of the base models' final predictions to improve the classification. Model voting techniques are already used to perform either average or majority votes to get final predictions. The other work to improve Covid-19 prediction used ensemble learning methods and Bagging techniques to train the base model on different instances of the training data and aggregate the model instances. Boosting techniques aggregate different instances of the model where each model instance is an improved version of the previous model [28]. In this paper, we take advantage of the feature maps generated by Deep CNN architectures rather than the final predictions produced by the model. These deep feature maps are concatenated and fed into machine learning or deep meta-learners. The intuition behind this approach is that the deep feature maps characterize generic information about the data hence allowing the meta-learners to extract better features pertaining to the data. [29] presented an adaptive boosting technique to improve ResNet model for CT scan inference. [30] studied a gradient boosting technique

over wide ResNet and compared its results to ResNet, DenseNet and InceptionNet. The boosted ensemble gave an accuracy of 87.7% for CT scan inference. [31] presented a multi stacked ensemble where the output of three meta learners namely SVM, Autoencoder, Naive Bayes were combined using model voting for the final output. [32] presented a voting Ensemble of three base CNNs where majority voting of the final probabilities gave the final Ensemble output. [33] studied stacking generalization ensemble of the VGG where the database was bootstrapped to improve the VGG classifier's accuracy to 93.57%.

Most of the existing Ensemble techniques present voting, stacking, boosting and / or bagging the final predictions of the base learners. The Ensemble study also limits to two or three CNNs as base learners and concentrates on binary classification which pertains less variance in the dataset. These models use Deep CNN classifiers as the base learners where the output feature maps from the Deep CNNs are passed onto a classification head to obtain the final probabilities. But the feature maps represent deep spatial information pertaining to the CT scan image which is lost when flattened for the classification head. Therefore, the feature maps offer to be a richer embedding compared to the final probabilities. This makes the feature maps more suited to be the input of the meta-learner while Ensemble modelling. The presented Ensemble models use the feature maps from four Deep CNN feature extractors and pass them to shallow and deep meta-learners for better generalization. The evaluation of these can be improved by using N-fold cross validation thereby utilizing the entire database.

Deep Learning approaches have been found to outperform learning machine algorithms while processing large databases with vast input features. With the introduction of attention mechanisms in [34], recent architectures like Transformers have extensively been used for state-of-the-art (SOTA) published results [35], [36]. Attention allows the model to focus on other embeddings in the ensemble that need to be paid attention to while processing one particular embedding which allows for a more direct dependence between the embeddings of the Ensemble. Convolutional Neural Networks have been traditionally used for extracting deep features from image datasets [37] and have been extended to other machine learning tasks due to their ability to capture complex and positionally invariant feature maps. From visualizations of the final convolutional activation representations of various Deep CNN architectures, it was observed that these models learn similar complex and positional features pertaining to the data [38], [39]. Hence, it would be reasonable to learn the inter-dependence between the Ensemble feature maps through attention.

This paper contributes to the development of different ensemble learning mechanisms for machine learning and CNN to extract deep features from the chest CT scan image by utilizing the concept of deep meta-learning and attention mechanism. Initially Deep CNN architectures namely VGG-19, Inception-V3, ResNet-152-v2, DenseNet-201, InceptionResNet-v2, Xception and Efficientnet-B7 that trained with five-fold cross-validation to extract deep features from the chest CT scan images. These classification models were optimized for the dataset by tuning their hyper-parameters and were then evaluated on the test image database. Out of these, the four best-performing models (ResNet-152-v2, Densenet-201, Xception, and EfficientNet-B7) have proceeded for ensemble learning. This paper proposes four groups of Ensemble learning mechanisms to predict COVID-19 from CT scan images.

- i. Normal machine learning meta-learning Ensemble CNNs models where the feature vectors extracted by the deep CNNs are concatenated and fed to various machine-learning models to classify the infected CT scan images.
- ii. Deep meta-learning Ensemble CNNs models that concatenate the extracted feature maps from the trained CNN models and then use Neural Network (NN) or Fully Convolutional Network (FCN) architectures to classify the infected CT scan images.

- iii. A novel Attentive Ensemble CNNs model that uses attention encoders to dynamically learn the relative importance between the feature maps extracted by the deep CNNs so as to classify the infected CT scan images.
- iv. Voting Ensemble Learning models that classify CT scan images by average voting or majority voting of the predicted probabilities of the four trained CNN models.

The rest of the paper is organized as follows. Section 2 provides a background of different deep CNN architectures followed by brief descriptions of various machine learning and deep learning models used as meta-learners for ensemble modelling. Section 3 explains the methodology of existing and proposed ensemble learning approaches. Experimental evaluation and results are presented in Section 4 and finally, Section 5 concludes the paper.

2. Materials and Methods

2.1. Database

This study utilized the dataset of COVID-19, Normal, and Pneumonia chest CT scan axial slices from [40–45] integrated in [46] after removing duplicated datasets. The dataset has been used in COVID-19 diagnosis literature and has proven its effectiveness and efficiency in deep learning applications.

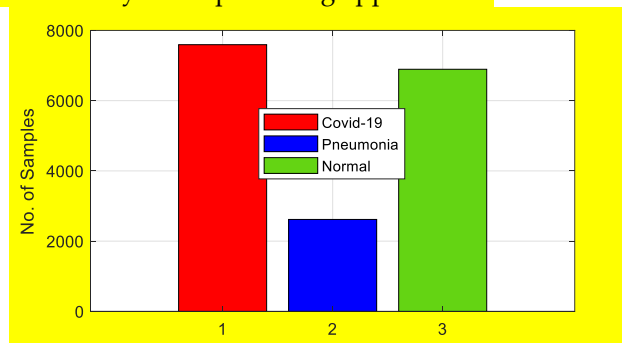


Figure 1. Distribution of Classes in the Image Database.

The entire database contains infected chest CT scan images of 7,593 COVID-19 cases and 2,618 viral pneumonia cases, as well as 6,893 normal cases as shown in Figure 1.





Figure 2. Covid 19 Sample CT scans.

366

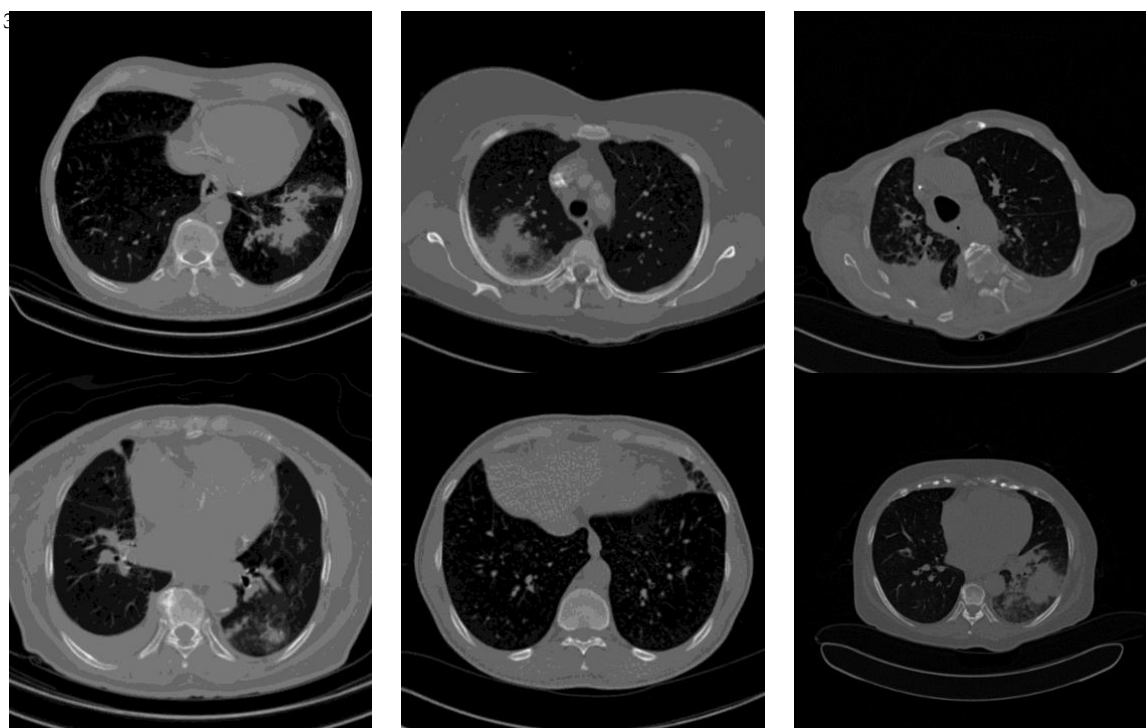
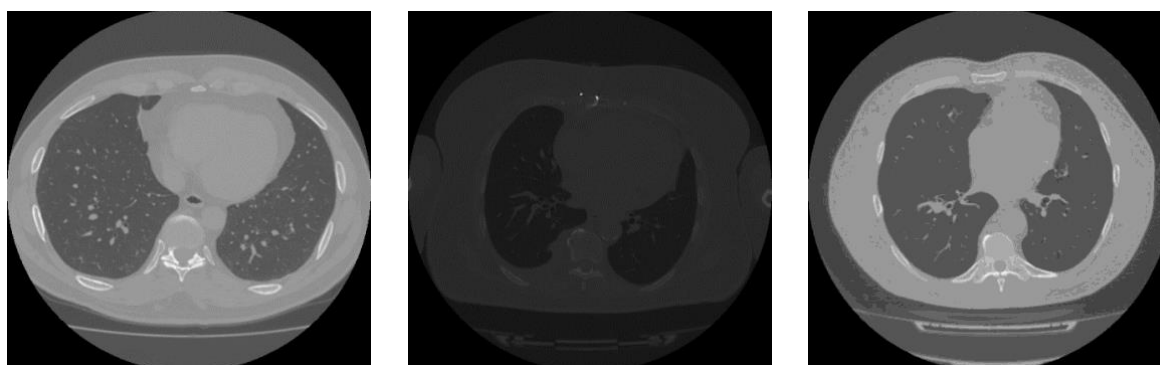


Figure 3. Viral Pneumonia Sample CT scans.

368



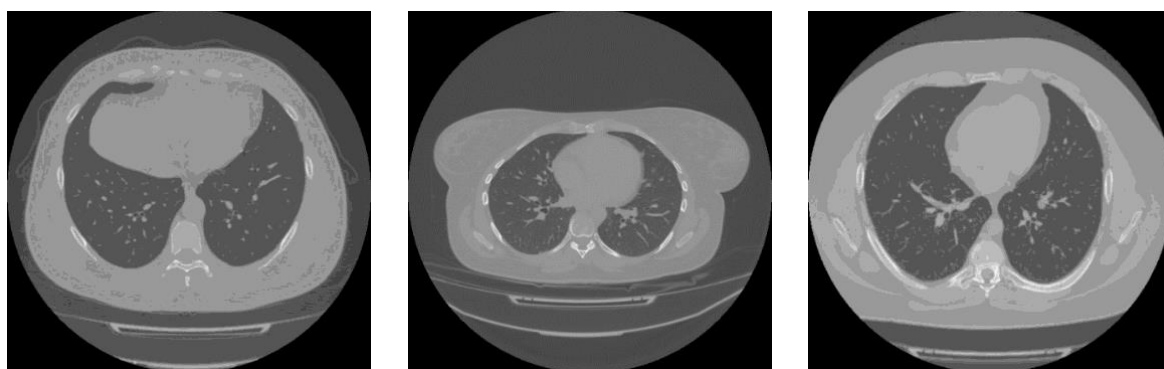


Figure 4. Normal Sample CT scans.

Figure 2 provides some sample lung CT scan images of COVID-19 infected patients. Figure 3 and Figure 4 shows some lung CT scan image samples of viral pneumonia patients and some normal lung CT scan image samples respectively.

2.2. Deep CNN Architectures

Convolutional Neural Networks (CNNs) have been traditionally used for extracting deep features from image datasets [47] and have been extended to other machine learning tasks due to its ability to capture complex and positionally invariant feature maps. CNNs train convolutional kernels to extract simple low-level features in shallow layers and complex high-level features in deeper layers. CNNs make use of local receptive fields, in which hidden units are connected to local patches of the lower layer. This allows the network to capture local spatial relationships of the image pixels. Weight sharing of the kernels enables translation-invariance of the network to objects in images. It also reduces the number of parameters of the network. CNNs also use pooling as a parameter free down sampling operation serving to aggregate information. This section briefly discusses different CNN architectures used in this work such as VGG-19, Inception-V3, ResNet-152-v2, InceptionResNet-V2, DenseNet-201, XceptionNet and EfficientNet-B7. The Deep CNNs automatically extract important features pertaining to the dataset thereby removing the need for handcrafted feature engineering.

VGG-19 [38] is a standard deep CNN with 19 convolutional layers. VGG-19 employs stacks of 3x3 convolutional kernels with max-pooling layers in between them. This produces a simple deep architecture where the 3x3 kernels extract complex features and the max-pooling layers amplify the important features.

The Inception architecture runs multiple parallel convolutions using various kernels to extract features at various scales [48] as opposed to typical CNNs where the convolutions are stacked sequentially. The architecture proposes a wider as well as deeper architecture. The third version of the Inception family is the 48-layer deep Inception-v3 which utilizes 1x1 pointwise convolution layers to reduce the feature dimensionality and parameters of the model as well as improve computational efficiency.

Resnet-152-v2 architecture is 152 layers deep. This version of the architecture performs batch normalization and ReLU activation before the convolution for a better identity mapping [49] and is employed in this paper.

The Inception-ResNet [50] architecture utilizes residual connections between the Inception blocks allowing for better gradient propagation thereby improving the Inception network's depth as well as training speed. Inception-ResNet-v2 which is a 164-layer deep architecture was utilized in this work.

The DenseNet [51] architecture aims to tackle the vanishing gradient problem in classical CNNs by introducing dense blocks where each convolution output is forward propagated to every other convolution layer through skip connections. This study used the DenseNet-201 architecture, that has 201 layers.

Xception (“Extreme version of Inception”) is a 71-layer deep Inception-inspired architecture that utilizes depth wise separable convolutions and 1x1 pointwise convolutions to substantially decrease the model parameters and improving computation efficiency thereby preventing the model from overfitting the dataset [52].

EfficientNet proposed a compound factor to consistently scale the baseline model’s parameters such as resolution, width and depth. This compound scaling allows the model to adapt to different image sizes by introducing additional layers or channels to extract deep features [53]. The baseline model is obtained by performing Neural Architecture Search on the dataset. EfficientNet-B7 inspired by the architecture searched for ImageNet dataset was utilized in this research.

Ensemble modelling is an approach where multiple separate models are developed to predict an outcome, either using a variety of modelling techniques or bagging the training database. The meta-learners then aggregate each base model’s predictions to provide a single final prediction. The purpose of employing Ensemble models is to reduce the generalization error on the test database by reducing the variance.

2.3. Deep Meta-Learners

Ensemble modelling had been used in various fields and found to produce enhanced results [54]. Deep-learning models used as meta-learners for Ensemble modelling are Neural Network (NN) Architecture, Fully Convolutional Network (FCN) Architecture and Multi-Head Attention Encoder. Since the former two have been thoroughly discussed in many articles, a brief overview of Multi-Head Attention Encoder alone is outlined as the paper purpose a novel approach based on the same.

The Transformer encoder is made of several stacked encoder modules each of which feeds into the encoder at the next level. Each encoder module employs sub-layers of Multi-Head Attention and feedforward network where each sub-layer adopts a residual connection and layer-normalization. The attention encoder output does not depend on order of inputs. Therefore, positional embeddings were added to the feature embeddings so that information about order of sequence is retained. The embedded sequence was passed as query (Q), key (K) and value (V) to the attention encoder module.

Self-Attention allows the model to focus on other embeddings in the sequence that need to be paid attention to while processing one particular embedding. Scaled dot-product attention mechanism was utilized in the self-attention layer as it obtains more stable gradients by scaling the values to a manageable range.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Multi-Head Attention concatenates multiple self-attention representations to get the attention output thereby expanding the model’s ability to focus on different parts of the sequence. The attention output is then passed on to the next encoder module through a feedforward network with residual connections between sublayers

Figure 5 illustrates the Multi-Head Attention Encoder architecture.

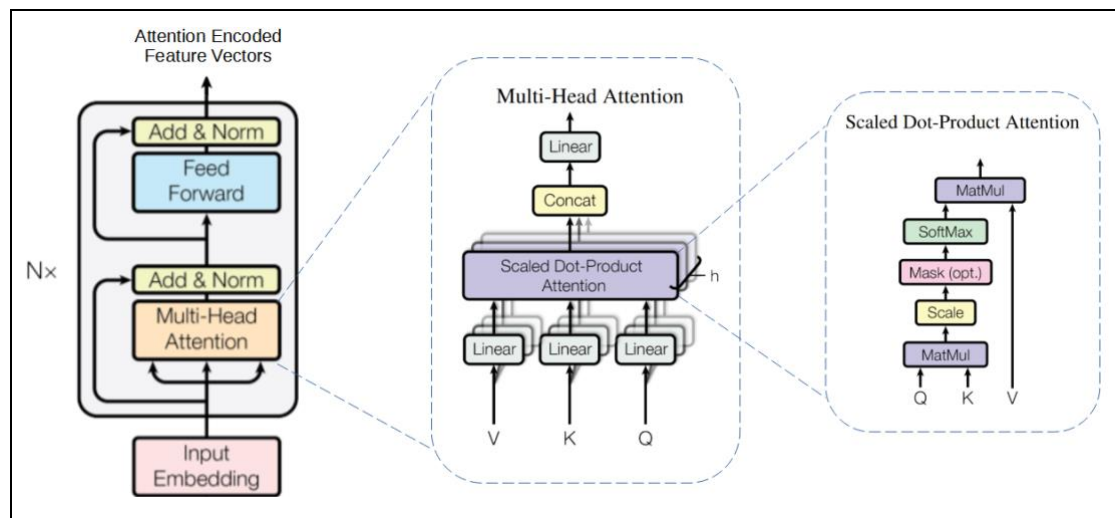


Figure 5. Multi-Head Attention Encoder.

2.4. Methodology

This section explains the proposed Ensemble Learning methodology illustrated in Figure 6. The proposed method consists of features extraction layer and a classification layer. Various Deep CNNs are used as feature extractors. The Ensemble models represent chest CT scan classifiers with Deep CNNs as feature-extracting base learners and the presented deep and shallow classifiers as meta-learners. Seven popular Deep CNN architectures were trained and evaluated on the CT scan database. The Deep CNNs are coupled with the densely connected network for classifying the chest CT scan images. These models are then trained and hyperparameter tuned to obtain the optimum Deep CNN classifiers. The four top performing models were used as trained base feature extractors for Ensemble learning. These Deep CNN feature extractors are used as base learners in the Ensemble. The features are in the form of model probabilities or deep feature vector or deep feature map. The classification ensemble models such as average ensemble, majority ensemble, shallow machine learning ensembles, NN ensembles, CNN ensembles and attention-based ensembles are experimented for synergizing the output features from the Deep CNNs and perform the final classification.

Various state-of-the-art Deep CNN architectures are used to extract deep feature representations from the CT scan image database. VGG-19, Inception-V3 and ResNet-152-v2, Densenet-201, InceptionResNet-V2, Xception and Efficientnet-B7 are used as feature extractors to obtain deep feature maps that were then used to train a fully connected neural network classification head to get the final model prediction. The hyperparameters of the Deep CNN feature extractors are optimized in [38, 48,49,51-53]. These architectures are initialized with ImageNet weights for transfer learning on 224x224 sized images. The classification head was hyper-parameter tuned with different numbers of layers, a different number of neurons per layer and dropout factors. Dropout regularization is used to prevent the classification head from overfitting the training fold database and generalize better on the validation fold. ReLU activation was employed for learning the non-linearity. The CT scan images were first preprocessed by resizing them to 224x224 pixels using Bilinear interpolation and rescaling the image intensity to the range [-1,1] so as to get it into the correct format pertaining to the architectures thereby making efficient utilization of transfer learning.

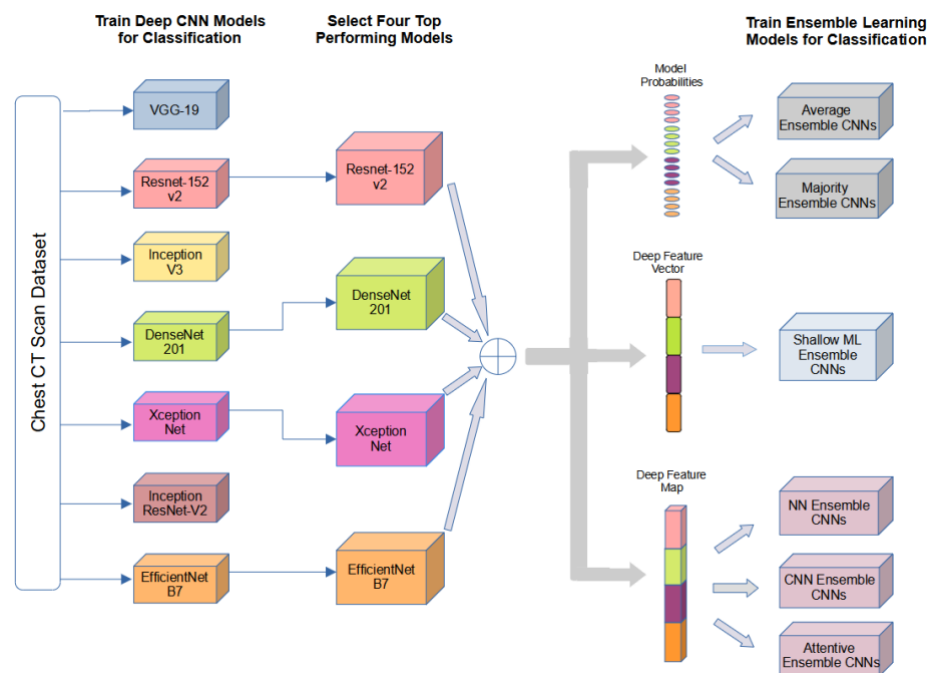


Figure 6. Overview of Methodology.

Average and Majority Ensemble CNNs are described in section 2.4.1. Shallow Ensemble CNNs are described in section 2.4.2. NN Ensemble CNNs are described in section 2.4.3. CNN Ensemble CNNs are described in section 2.4.4. Attentive Ensemble CNNs are described in section 2.4.5.

A four-layer neural network classification head with 25% dropout regularization gave best results for VGG-19, Inception-v3, ResNet-152-v2, InceptionResNet-v2 models while a three-layer neural network classification head with 50% dropout regularization gave best results for Densenet-201, Efficientnet-B7 and Xception models. The models were trained using Adam [5656] as an optimizer with a learning rate of 0.001 and the calculated loss being categorical cross-entropy. Figure 7 demonstrates training a Deep CNN model for feature extraction. The top four best-performing architectures obtained on evaluating the test was saved as base models for future Ensemble Deep CNNs model.

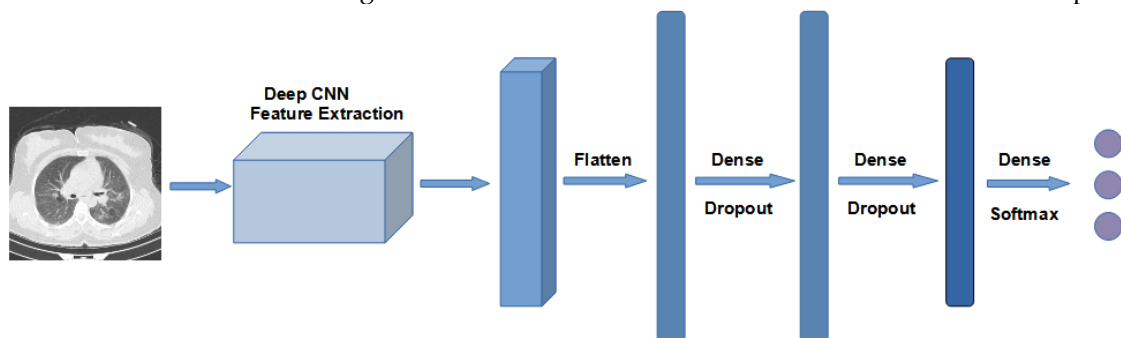


Figure 7. Training Deep CNN Models.

2.4.1. Deep Voting Ensemble CNN Models

In the Voting Ensemble CNN Model, as shown in Figure 8, the softmax scores obtained from the four trained classification models corresponding to each class were maxed or averaged to obtain the Ensemble model's predicted class. These models do not require additional training on 5-fold cross-validation.

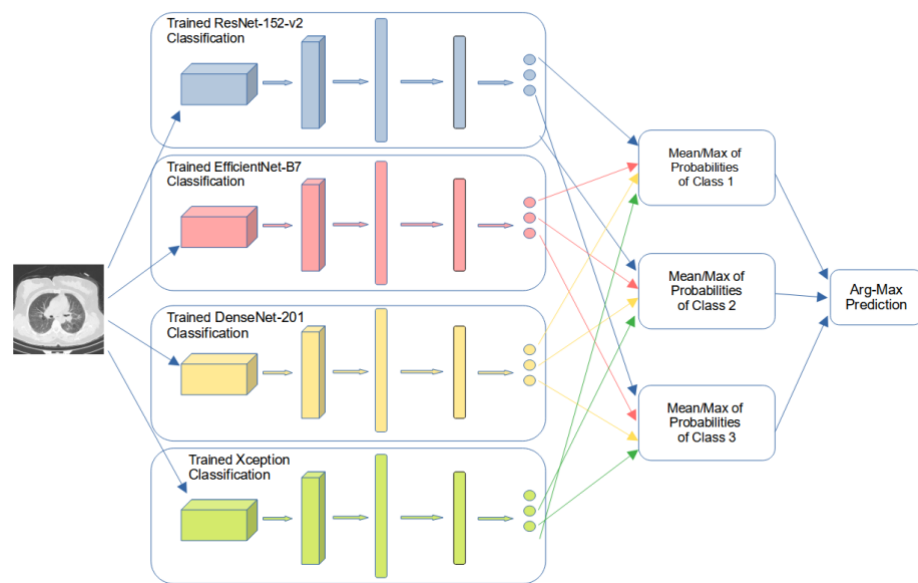


Figure 8. Deep Voting Ensemble Model.

2.4.2. Shallow Meta-Learning Ensemble CNN Models

For Deep Ensemble CNNs using Shallow Meta-Learners, the feature vectors obtained from the trained deep CNN feature extractors were first stacked to get a deep feature vector. This deep feature vector was then fed to a machine learning classifier. Logistic Regression, Support Vector Machine, Decision Tree, Gaussian Naïve Bayes, K-Nearest-Neighbour (KNN), Random Forest Ensemble, Bagging Ensemble, AdaBoost Ensemble and Gradient Boost Ensemble models were leveraged as meta-learning classifiers for the deep ensemble approach. The Logistic Ensemble CNNs Model was trained with a regularizing strength of 40 and l2 penalization. The model was trained for 100 iterations for convergence. The SVC Ensemble CNNs Model gave best results using RBF kernel, regularizing strength of 40 and l2 penalization. The Decision Tree Ensemble CNNs Model was hyper parameterized using Gini impurity criterion for splits and the tree was allowed to deepen till all leaves are pure or has 2 samples. The Naïve Bayes Ensemble CNNs Model used the Gaussian Naïve Bayes algorithm for classification. The KNN Ensemble CNNs Model was trained with K = 10 obtained using elbow method. The Random Forest Ensemble CNNs Model was tuned to 100 tree estimators without bootstrapping dataset and Gini impurity as the split condition to get optimum results. Bagging Ensemble CNNs Model using 30 Decision tree base estimators and bootstrapped dataset gave the best result on test database. The AdaBoost Ensemble Model with 100 Decision Tree base learners and a learning rate of 0.7 gave optimum results. Gradient Boosting Ensemble Model was hyper parameterized with 100 base estimator boosting rounds with a learning rate of 0.9 for best results. Figure 9 depicts the proposed approach.

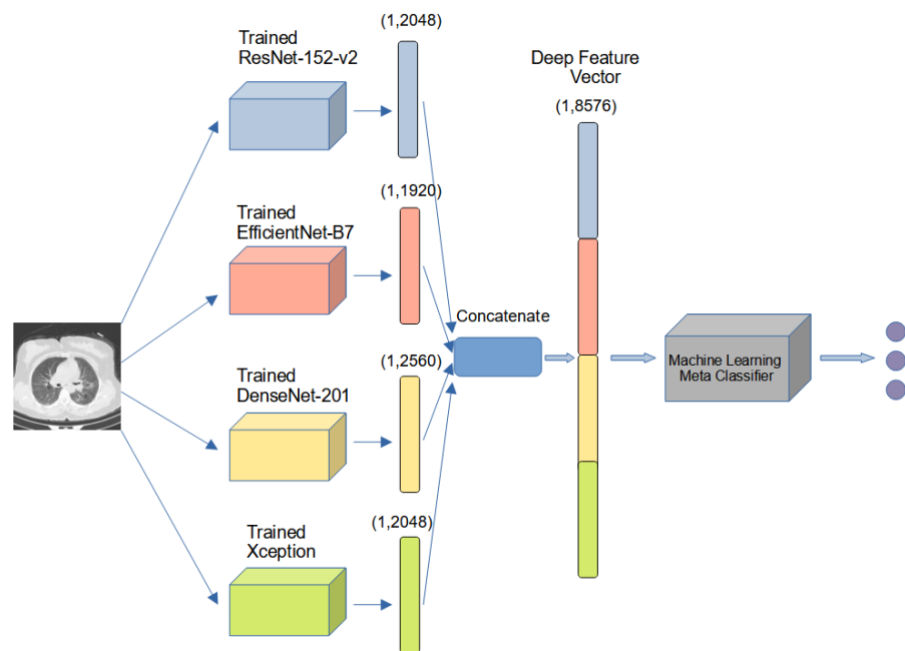


Figure 9. Shallow Meta-Learning Ensemble CNN Model.

2.4.3. NN Ensemble CNN Model

In the NN Ensemble CNN meta-learning approach [57], the feature maps obtained from the final convolutional layer of the four trained Deep CNN models were global average pooled to get deep feature vectors. Global Average Pooling takes the average of each feature map thereby reducing the size of the features and the parameters to get a better representational mapping to the class labels. These four feature vectors were then combined to get the Ensemble deep feature vectors.

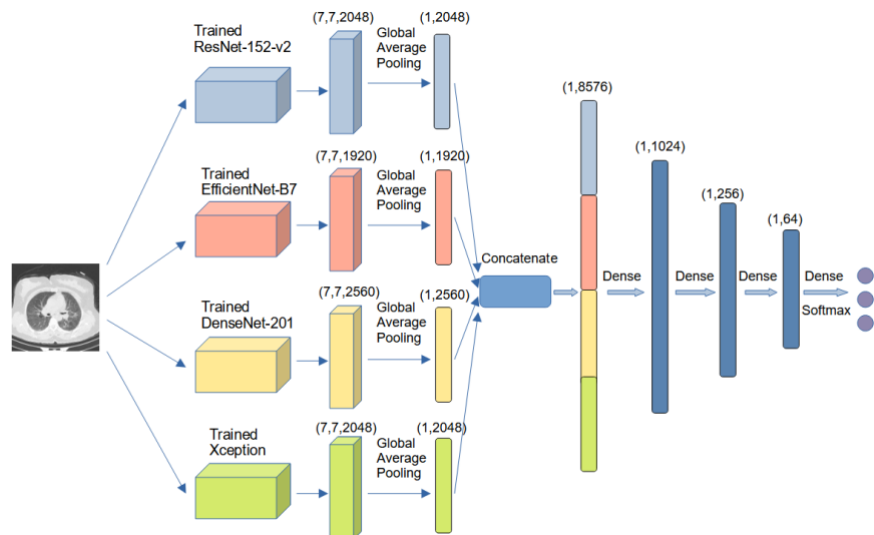


Figure 10. NN Ensemble CNN Model.

A feed-forward neural network (NN) classification head was then employed to meta-learn the deep ensemble feature vector along with the final softmax activation which gives the class probabilities. The NN classification head was hyper-parameter tuned with different numbers of layers, number of neurons per layer and dropout factors. After tuning, the best validation accuracy obtained was by using a four-layer NN with 1024, 256, 64 and 3 neurons respectively to each layer and 0.2% dropout, which

is shown in Figure 10. The maximum of the softmax scores was used as the Ensemble model's predicted class for classification.

2.4.4. FCN Ensemble CNN Model

In the FCN Ensemble CNN meta-learning approach [57], feature map outputs from the final convolutional layer of the four trained Deep CNN models were then concatenated along the channels to obtain the Ensemble deep feature map. A Fully Convolutional Network (FCN) was then meta-learned to get the final softmax predictions. The FCN architecture was hyper-parameter tuned with different convolutional layers and different numbers of various convolutional and max-pooling kernel sizes (2x2, 3x3 and 4x4). After experimenting, an FCN modelled with three convolutional blocks made of 2x2 convolutional and max-pooling kernels of 2048, 512, and 64 kernels per block respectively gave the best result on the validation database. Figure 11 illustrates the proposed approach. The final feature map was then flattened and passed through a dense layer with softmax activation to get the final classification scores. The maximum of the softmax scores was used as the Ensemble model's predicted class for classification.

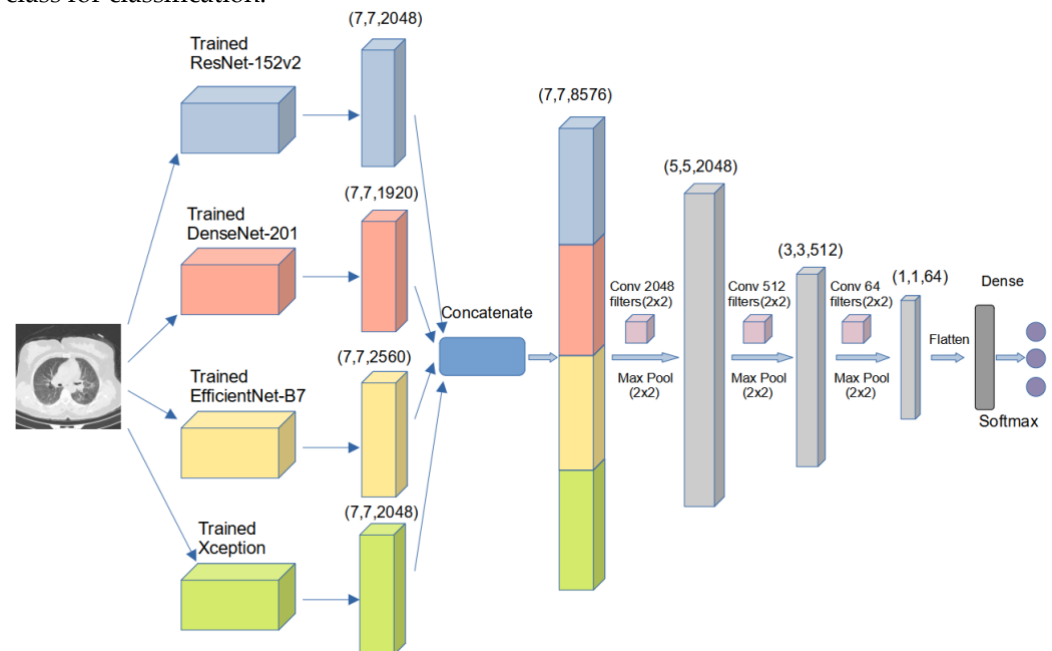


Figure 11. FCN Ensemble CNN Model.

2.4.5. Attentive Ensemble CNN Model

In the Attentive Ensemble-CNN model, the feature maps corresponding to the final convolutional layers were extracted from the four trained models. 1x1 pointwise convolution was performed on each of these feature maps to reduce the channels of the maps to the same depth. These maps were then passed as input feature embedding to three stacks of Multi-Head Attention encoders. The encoder was constructed using a multi-head attention layer that used a scaled dot product attention mechanism and a Feed Forward Network made of 3x3 convolution layer. Skip connections were also employed between the sub-layers for better feature propagation. Each of these encoders learns the relative importance of the feature embeddings with respect to the other embeddings. The attention-encoded feature maps are then passed onto the next encoder in a feed-forward fashion as shown in Figure 12. The final encoded feature maps were then concatenated and passed through a dense layer with softmax activation to get the final classification scores. The maximum of the softmax scores assigned the Ensemble model's predicted class for classification.

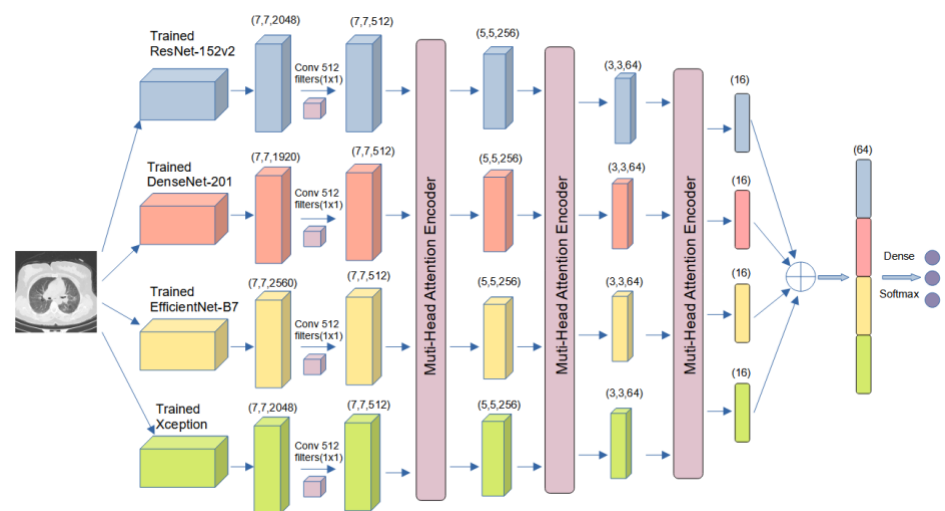


Figure 12. Attentive Ensemble CNN Model.

2.5. Experiment Setup

The Deep CNN models and the Ensemble learning models were built, trained, and evaluated using the TensorFlow / Keras framework with Python programming on cloud TPU engine (8 TPU v2 cores with 64GB total memory performing 180 TFlops). The image database was split into five-fold cross validation and test databases as shown in Table 1 using a pseudo random generator and loaded into the TPU cache memory. 95% of the database was used for 5-fold cross validation. The rest 5% of the database was used for testing. In the 5-fold cross validation dataset, 80% was used as training and 20% was used for cross validation during each fold. Each model was trained using five-fold cross validation before being restored to the epoch with the best results on the validation database. To exploit TPU performance, images with a resolution of 224x224 pixels were sent to the models in mini-batches of 80.

Table 1. Fivefold cross validation and testing CT scan image distribution.

CT Scan Database Split		Count
5-Fold Cross Validation	Training	13000
	Validation	3248
	Testing	856

3. Results

3.1. Evaluating Single Deep CNN Architectures

Adam optimizer [56] with a learning rate of 0.001 was utilized to train the Deep CNN classification models using categorical cross-entropy as the computed loss. The evaluation metrics obtained by these models on the test database are shown in Table 2. It is evident from the table that ResNet-152-v2, Densenet-201, Efficientnet-B7 and Xception architectures gave better accuracy and average F1-scores on the test database as compared to VGG-19, Inception-v3 and InceptionResNet-v2. These four best performing models were saved and used as base models for future Ensemble learning approaches. The Xception Model gave the best accuracy of 99.18% while the VGG-19 model gave the worst accuracy of 96.18%. All the models except VGG-19 were able to classify the CT scan images with accuracy and F1-score more than 98.65%. The Inception-v3, ResNet-152-v2, Efficientnet-B7 and Xception models were able to classify the Viral Pneumonia CT scan images with a perfect score of 1. The DenseNet-201,

ResNet-152-v2, Xception and EfficientNet-b7 models performed fairly even at high thresholds by producing an AUC value more than 99.2%.

Table 2. Evaluation Metrics for Deep CNN Models.

	Accuracy	Class Label	Precision	Recall	F1 Score	AUC
VGG-19 Model	0.9678	COVID-19	0.9810	0.9453	0.9628	0.9585
		Normal	0.9430	0.9821	0.9622	0.9525
		Viral	0.9926	0.9926	0.9926	0.9956
		Pneumonia				
Inception-v3 Model	0.9871	COVID-19	0.9946	0.9765	0.9855	0.9879
		Normal	0.9738	0.9940	0.9838	0.9849
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
ResNet-152-v2 Model	0.9895	COVID-19	0.9921	0.9843	0.9882	0.9897
		Normal	0.9823	0.9910	0.9867	0.9882
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
Inception ResNet-v2 Model	0.9883	COVID-19	1.0000	0.9765	0.9881	0.9906
		Normal	0.9795	0.9970	0.9882	0.9888
		Viral	0.9782	1.0000	0.9890	0.9891
		Pneumonia				
DenseNet-201 Model	0.9884	COVID-19	0.9794	0.9947	0.9870	0.9875
		Normal	0.9945	0.9792	0.9871	0.9903
		Viral	1.0000	0.9925	0.9962	0.9993
		Pneumonia				
Xception Model	0.9918	COVID-19	0.9947	0.9870	0.9908	0.9921
		Normal	0.9852	0.9940	0.9896	0.9907
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
EfficientNet-b7 Model	0.9895	COVID-19	0.9947	0.9817	0.9867	0.9900
		Normal	0.9795	0.9940	0.9867	0.9878
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				

3.2. Evaluating Voting Ensemble CNN Models

For the Average and Majority Voting Ensemble models, the four saved trained models (ResNet-152-v2, Densenet-201, Xception and Efficientnet-b7) were used as base models. The Voting Ensemble model, evaluated on the test database produced the evaluation metrics as shown in Table 3. Both the models produced an accuracy and average F1 score more than 99.4% thereby outperforming the deep CNN architectures. Both the models performed relatively well even at high thresholds by producing an AUC value more than 99.56% outperforming the Deep CNN models.

Table 3. Evaluation Metrics for Voting Ensemble CNN Models.

	Accuracy	Class Label	Precision	Recall	F1 Score	AUC
Average Ensemble CNN Model	0.9953	COVID-19	1.0000	0.9895	0.9947	0.9957
		Normal	0.9882	1.0000	0.9941	0.9941
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
Majority Ensemble CNN Model	0.9942	COVID-19	1.0000	0.9869	0.9934	0.9947
		Normal	0.9853	1.0000	0.9926	0.9926
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				

3.3. Evaluating Shallow Meta-Learning Ensemble CNN Models

For the Shallow meta-learning Ensemble CNNs models, the four trained Deep CNN feature extractors were used as ensemble base models to train the machine learning meta-learners as shown in Figure 5. Table 4 shows the metrics obtained after evaluating the models on the test database. The Logistic, KNN, Random Forest, Bagging, AdaBoost and Gradient Boosting meta learners achieved accuracy and an average F1 score of more than 99.18% thereby outperforming the deep CNN architectures. Random Forest and Gradient Boosting Ensemble CNNs models produced the best performance but were not able to outperform the Voting Ensemble CNNs models. The KNN, Random Forest, AdaBoost, Gradient Boosting and Bagging Ensemble CNNs models performed relatively well even at high thresholds by producing an AUC value more than 99.5% outperforming the Deep CNN models. The rest of the shallow meta-learners performed relatively poor at higher classification thresholds. The Shallow meta-learning Ensemble CNNs models were not able to outperform the Voting Ensemble CNNs models in terms of AUC.

Table 4. Evaluation Metrics for Deep Ensemble Models using ML classifiers.

	Accuracy	Class Label	Precision	Recall	F1 Score	AUC
Logistic Ensemble CNNs Model	0.9918	COVID-19	0.9896	0.9921	0.9908	0.9916
		Normal	0.9910	0.9881	0.9895	0.9916
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
SVC Ensemble CNNs Model	0.9885	COVID-19	0.9795	0.9973	0.9883	0.9874
		Normal	0.9969	0.9762	0.9865	0.9899
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
Decision Tree Ensemble CNNs Model	0.9907	COVID-19	0.9947	0.9843	0.9895	0.9910
		Normal	0.9824	0.9940	0.9882	0.9892
		Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
Naïve Bayes Ensemble CNNs Model	0.9883	COVID-19	0.9794	0.9947	0.9870	0.9875
		Normal	0.9939	0.9792	0.9865	0.9903
		Viral	1.0000	0.9925	0.9962	0.9993
		Pneumonia				
KNN Ensemble CNNs Model	0.9929	COVID-19	0.9922	0.9947	0.9934	0.9939
		Normal	0.9940	0.9881	0.9910	0.9931
		Viral	0.9926	1.0000	0.9963	0.9963
		Pneumonia				

Random		COVID-19	0.9911	0.9970	0.9940	0.9955
Forest		Normal	0.9973	0.9921	0.9947	0.994
Ensemble	0.9953	Viral	1.0000	1.0000	1.0000	1
CNNs Model		Pneumonia				
Bagging		COVID-19	0.9947	0.9869	0.9908	0.9921
Ensemble		Normal	0.9852	0.9940	0.9896	0.9907
CNNs Model	0.9919	Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
AdaBoost		COVID-19	0.9973	0.9895	0.9934	0.9944
Ensemble		Normal	0.9882	0.9970	0.9926	0.9931
CNNs Model	0.9941	Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
Gradient		COVID-19	1.0000	0.9895	0.9947	0.9941
Boosting		Normal	0.9882	1.0000	0.9941	0.9957
Ensemble		Viral	1.0000	1.0000	1.0000	1
CNNs Model	0.9952	Pneumonia				

3.4. Evaluating Deep Meta-Learning Ensemble CNN Models

For the proposed Deep meta-learning Ensemble CNNs models, the four trained Deep CNN feature extractors were used as ensemble base models to train the deep learning architectures using hyperparameters as shown in Figure 6, Figure 7 and Figure 8. Adam optimizer [56] with a learning rate of 0.001 was utilized to train the ensemble classification models using categorical cross-entropy as the computed loss. Table 5 shows the evaluation metrics obtained after evaluating the models on the test database. The deep meta learners produced an accuracy and average F1 score more than 99.65% thereby outperforming the Deep CNN models, Voting Ensemble CNNs models and Shallow meta-learning Ensemble CNNs models. The Attentive Ensemble model achieved the best performance with an accuracy of 99.88% and classifying COVID-19 with an F1-score of 99.87.

Table 5. Evaluation Metrics for Deep Ensemble Models using deep learning classifiers.

	Accuracy	Class Label	Precision	Recall	F1 Score	AUC
NN		COVID-19	0.9948	0.9973	0.9960	0.9963
Ensemble		Normal	0.9970	0.9940	0.9955	0.9965
CNNs Model	0.9965	Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
FCN		COVID-19	1.0000	0.9948	0.9973	0.9978
Ensemble		Normal	0.9941	1.0000	0.9970	0.9970
CNNs Model	0.9977	Viral	1.0000	1.0000	1.0000	1
		Pneumonia				
Attentive		COVID-19	0.9974	1.0000	0.9987	0.9987
Ensemble		Normal	1.0000	0.9970	0.9985	0.9990
CNNs Model	0.9988	Viral	1.0000	1.0000	1.0000	1s
		Pneumonia				

3.5. Confusion Matrix for Deep CNN Models

Figure 13(a) – (g) depicts the confusion matrices obtained on evaluating the Deep CNN models on the test database. The DenseNet-201 and Xception models performed fairly on classifying the COVID-19 lung CT scans by misclassifying less than five. The

Inception-v3, InceptionResNet-v2, Xception and EfficientNet-b7 models performed fairly on classifying the normal lung CT scans by misclassifying less than three.

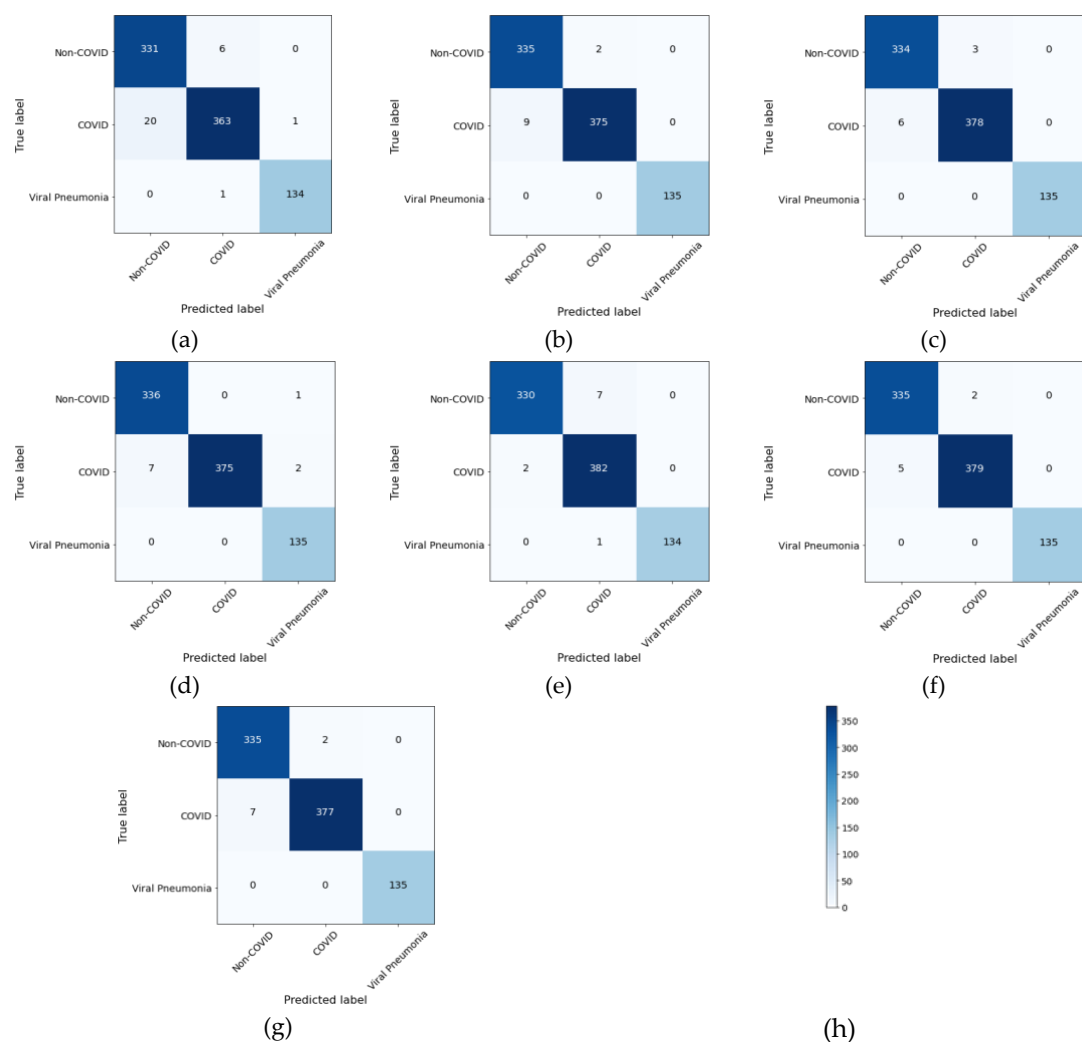


Figure 13. Confusion Matrix for Deep CNN Models (a) VGG-19 Model (b) Inception-v3 Model (c) ResNet-152-v2 Model (d) InceptionResNet-v2 Model (e) DenseNet-201 Model (f) Xception Model (g) EfficientNet-b7 Model. (h) Label

3.6. Confusion Matrix for Voting Ensemble Models

Figure 14(a)-(b) shows the confusion matrices obtained on evaluating the Voting Ensemble CNNs models on the test database. Both models performed well by perfectly classifying the normal and viral pneumonia lung CT scans and misclassifying less than 6 COVID-19 lung CT scans.

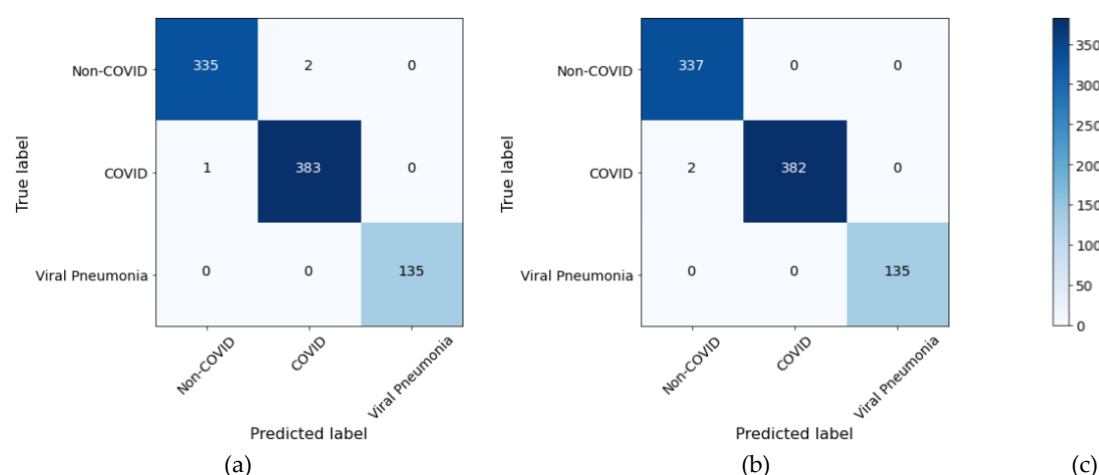
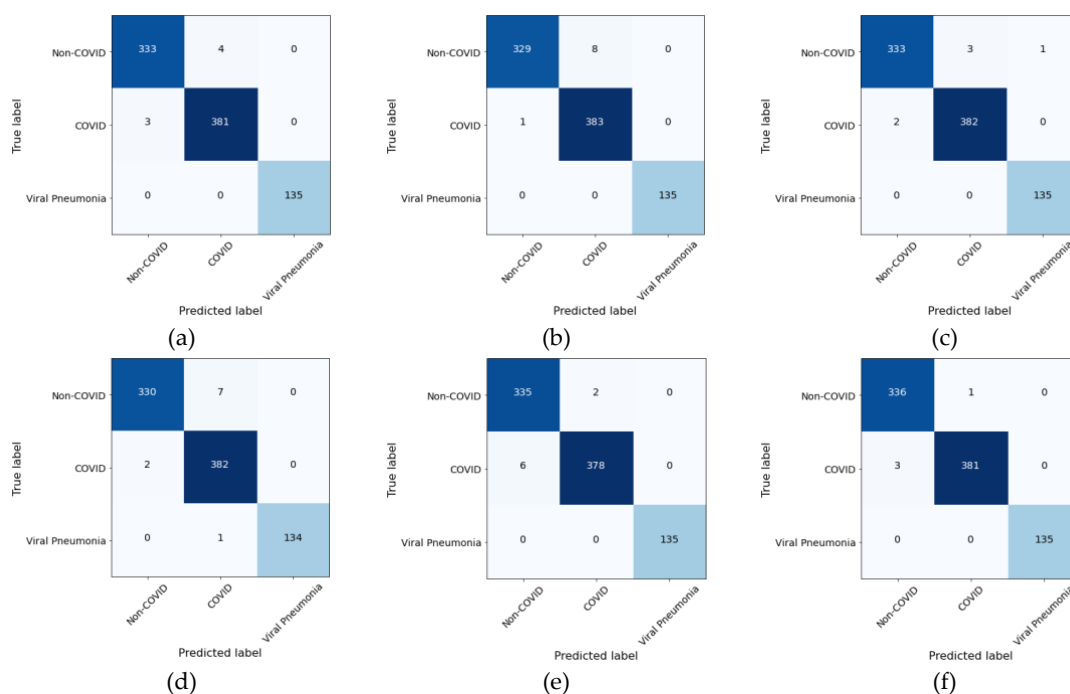


Figure 14. Confusion Matrix for (a) Average Ensemble CNN Model (b) Majority Ensemble CNN Model. (c) Label

3.7. Confusion Matrix for Shallow Meta-Learning Ensemble CNN Models

Figure 15(a) – (i) shows the confusion matrices obtained on evaluating Shallow meta-learning Ensemble CNNs models on the test database. The SVC, KNN, Naive Bayes and Random Forest Ensemble CNNs models performed really well on classifying the COVID-19 lung CT scans by misclassifying less than four. The Decision Tree, Random Forest, AdaBoost, Gradient Boosting and Bagging Ensemble CNNs models performed well in classifying the normal lung CT scans by misclassifying less than three. All these models misclassified one or zero viral pneumonia lung CT scans.



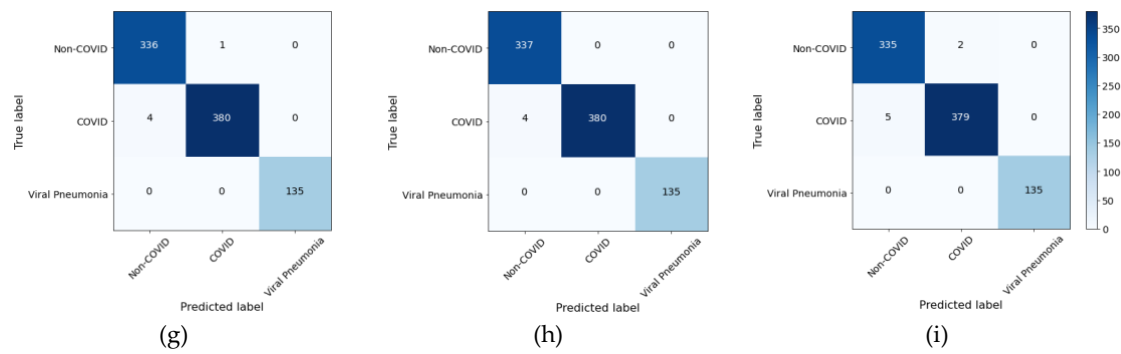


Figure 15. Confusion Matrix for Shallow Meta-Learning Ensemble CNN Models (a) Logistic (b) SVC (c) KNN (d) Naïve Bayes (e) Decision Tree (f) Random Forest (g) AdaBoost (h) Gradient Boosting (i) Bagging Ensemble CNN Model.

3.8. Confusion Matrix for Deep Meta-Learning Ensemble CNN Models

Figure 16(a)–(c) shows the confusion matrices obtained on evaluating the Deep meta-learning Ensemble CNNs models on the test database. From the confusion matrices it is evident that these models achieved the best performance by misclassifying less than 4. The Attentive Ensemble CNNs model perfectly classified all the test COVID-19 lung CT scan images while the FCN Ensemble CNNs model perfectly classified all the test normal lung CT scan images. All the models perfectly classified all the viral pneumonia test lung CT scans. The models produced the best results even at high thresholds by achieving an AUC value more than 99.75% outperforming the Deep CNN models and Shallow meta-learning Ensemble CNNs models.

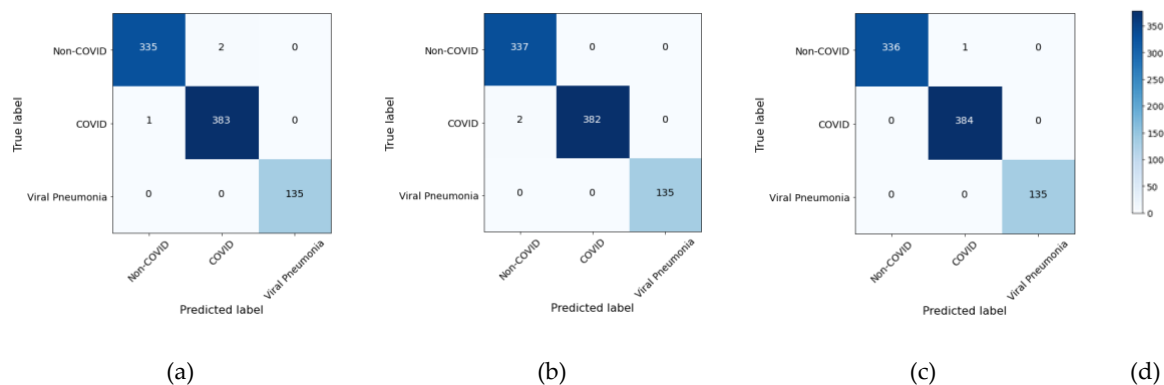


Figure 16. Confusion Matrix for Deep Meta-Learning Ensemble CNN Models (a) NN (b) FCN (c) Attentive Ensemble CNN Model. (d) Label

3.9. Robustness on Chinese CT Scan Database

The robustness of the proposed framework to generalization was tested by evaluating the trained model on a subset of a Chinese CT Scan database [58]. 1470 CT scan images were tested to classify them as COVID infection, common pneumonia or Normal condition. Table 6 shows the accuracy and F1 score of the evaluated models on Chinese CT Scan database. The Deep CNN models struggled to classify some images due to the variance in the Chinese data distribution compared to training database. The proposed models were able to perform decently thereby ensuring the generalization capability of the Ensemble approach.

Table 6. Evaluation Metrics on Chinese CT Scan database

Model	Accuracy	F1 Score
VGG-19	0.8880	0.8879
Inception-v3	0.8977	0.8974
ResNet-152-v2	0.9253	0.9251
InceptionResNet-v2	0.9499	0.9498
DenseNet-201	0.9482	0.9481
Xception	0.9511	0.9508
EfficientNet-b7	0.8794	0.8794
Average Ensemble CNNs	0.9641	0.9641
Majority Ensemble CNNs	0.9710	0.9698
Logistic Ensemble CNNs	0.9402	0.9401
SVC Ensemble CNNs	0.9395	0.9394
Decision Tree Ensemble CNNs	0.9513	0.9511
Naïve Bayes Ensemble CNNs	0.9381	0.9369
KNN Ensemble CNNs	0.9290	0.9287
Random Forest Ensemble CNNs	0.9644	0.9643
Bagging Ensemble CNNs	0.9597	0.9597
AdaBoost Ensemble CNNs	0.9644	0.9643
Gradient Boosting Ensemble CNNs	0.9750	0.9747
NN Ensemble CNNs	0.9766	0.9765
FCN Ensemble CNNs	0.9749	0.9749
Attentive Ensemble CNNs	0.9784	0.9782

3.10. Model Complexities

Table 7 shows the total parameters and the inference time (at 180 TFlops) of each of the presented Deep Learning models. The time complexity of the proposed approach includes time for training four base CNNs as well as training the meta-learners. The space complexity of the frameworks includes storing the CT scan images as input for Deep CNNs and also the output deep feature maps which will then serve as input for the meta learners. The trained parameters of the base learners and the meta learners need to be saved for inference. The hyper-parameters of the Deep CNN models are as proposed in works [48–53]. The Voting Ensemble CNNs models used the parameters of the entire Deep CNN classification models for final prediction while the presented Deep and Shallow Meta-Learning Ensemble CNNs models utilized the parameters of the Deep CNN feature extractors for meta-learning. The inference time was obtained after warning the TPU for few epochs on the test database. The Voting Ensemble CNNs Models used the least inference time while the Deep meta-learning Ensemble CNNs models consumed the most inference time.

Table 7. Model Complexities.

Model	Abbreviation	Model Type	Total Parameters	Inference Time (ms/image)
VGG-19	VGG	CNN	46,273,347	10.531
Inception-v3	IV3		74,790,435	6.062
ResNet-152-v2	R152		161,650,947	6.710
InceptionResNet-v2	IRV2		94,217,187	9.500
DenseNet-201	D201		115,218,755	11.34
Xception	XCEP		124,180,779	7.093

EfficientNet-b7	ENB7		193,107,098	10.718
Average Ensemble CNNs	AEC	Voting Ensemble	413,133,967	4.211
Majority Ensemble CNNs	MEC		413,133,967	4.208
Logistic Ensemble CNNs	LEC	Shallow Ensemble	161,627,799	4.615
SVC Ensemble CNNs	SVC		161,628,476	4.887
Decision Tree Ensemble CNNs	DTE		161,620,824	7.231
Naïve Bayes Ensemble CNNs	NBE		161,620,015	4.712
KNN Ensemble CNNs	KNN		161,620,324	4.698
Random Forest Ensemble CNNs	RFEC		161,621,950	7.723
Bagging Ensemble CNNs	BEC		161,624,916	8.072
AdaBoost Ensemble CNNs	ADA		161,624,377	8.588
Gradient Boosting Ensemble CNNs	GBEC		161,625,412	8.974
NN Ensemble CNNs	NNEC	Deep Ensemble	170,708,994	13.215
FCN Ensemble CNNs	FCNEC		236,229,890	14.327
Attentive Ensemble CNNs	ATEC		174,006,978	14.671

734

4. Discussion

735

Ensemble modelling combined the four trained deep CNNs allowing the meta-learners to decide predictions based on different features learned by the base models. The deep features are extracted from four trained models thereby ensuring independence between the features while Ensemble modelling. Comparing Table 3, Table 4 and Table 5 with Table 2, it is evident that the presented Deep Ensemble CNNs models outperform the deep CNN architectures portraying better accuracy and F1-scores. In contrast to the Conventional Ensemble Learning Models that aggregate the final predictions from the base models to produce the ensemble prediction, the Deep Ensemble Learning Models use deep feature vectors or feature maps from the trained models to train shallow or deep meta-learners. From the experiments, it can be deduced that using an Ensemble model with a strong meta-classifier like NN, CNN or Attention can outperform the Conventional and Shallow Ensemble models. The improved model complexity also allows the model to fit better on the training data thereby reducing the bias. By synergizing the deep feature maps, the Deep Ensemble models learn broad independent features pertaining to the data thereby allowing the models to generalize better on the test database.

Figure 21 and Figure 22 summarizes the prediction accuracies and COVID-19 classification F1-scores, obtained after evaluating the models on the test database. Out of the ensemble approaches, the proposed Attentive Ensemble CNNs Model produced better generalization in different evaluation metrics as it allowed the model to learn the relative significance between the ensemble feature embeddings.

Figure 23 and Figure 24 summarizes the total parameters and inference time of the presented models. The depth wise separable convolution in Xception feature extractor allows for lesser number of parameters thereby reducing overfitting. The dense block in DenseNet feature extractor allows for better gradient propagation and feature sharing. The Voting Ensemble CNNs models combined the four Deep CNN classification models where the classification heads took up the greatest number of parameters. The Voting models used the trained base models for inference and hence took less time for inference. The presented Deep and Shallow Meta-Learning Ensemble CNNs models employed meta-learners after the convolutional layers and hence contained lesser number of parameters. The Attentive Ensemble took the longest time for inference due to the computationally heavy attention mechanism.

The trained models can be directly served on a cloud service like Vertex AI on Google Cloud. The chest CT scan images obtained can be sent to the model endpoint using REST API provided by the cloud service provider. The model predictions can be

770

sent back to the end user thus aiding the medical diagnosis. The images can be sent in batches or in streams for offline or online predictions respectively depending on the application needs.

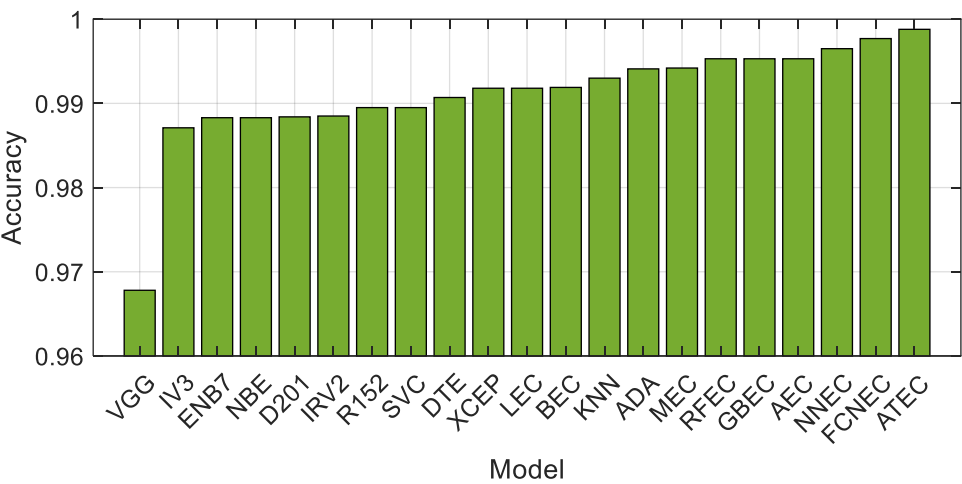


Figure 21. Prediction Accuracy.

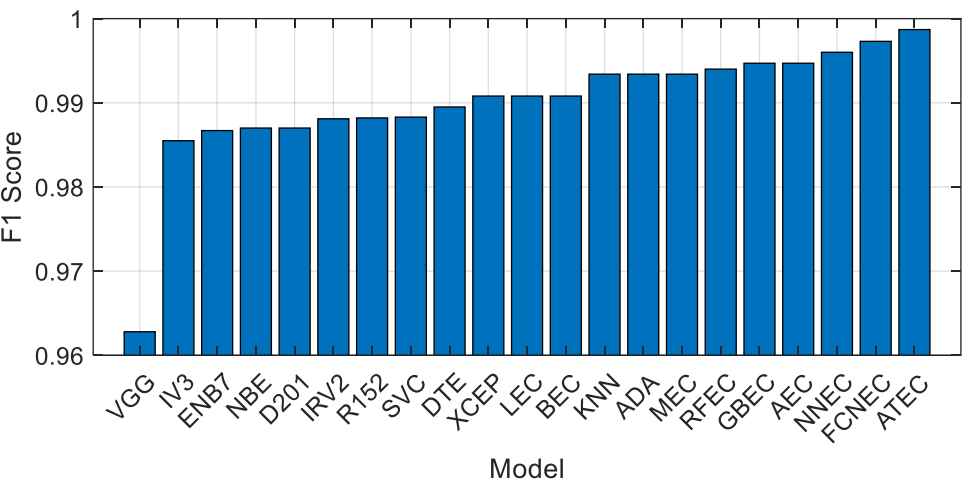


Figure 22. Prediction F1 Score.

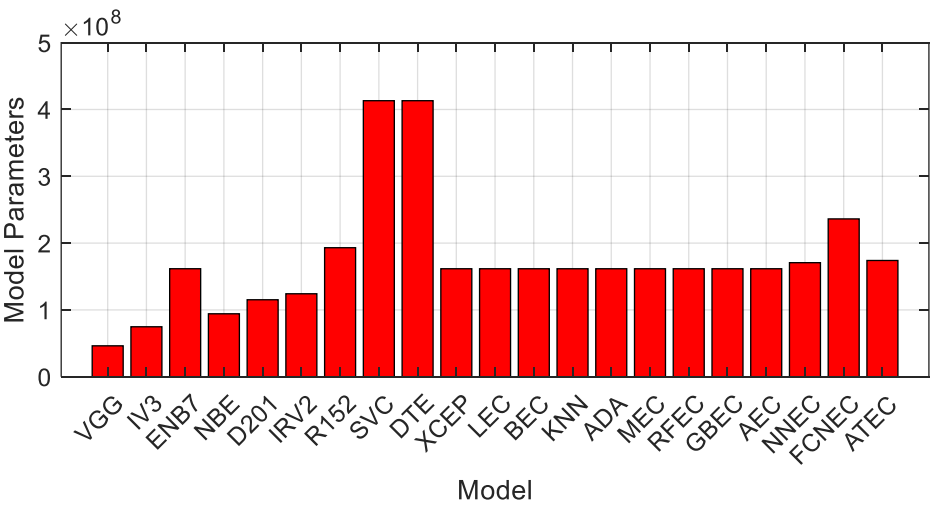


Figure 23. Ensemble Model Parameters.

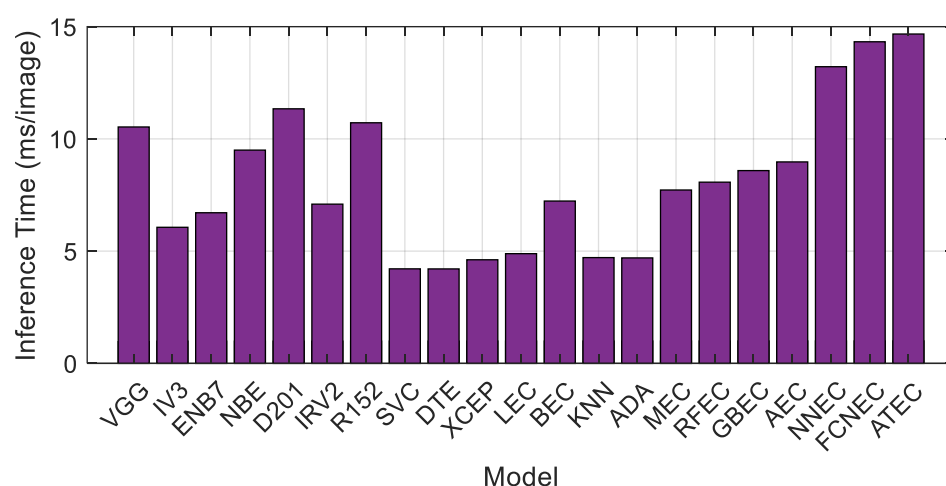


Figure 24. Inference Time of Ensemble Models.

5. Conclusions

This paper presented Deep Ensemble Learning architectures to classify infected chest CT scan images as COVID-19, as Viral Pneumonia or as Normal. Popular Deep CNN architectures were trained to classify these images using 5-fold cross-validation and the top four best performing models (ResNet-152-v2, Densenet-201, Xception and EfficientNet-B7) on the test database were employed as trained base models for Ensemble learning. Voting Ensemble CNNs models, Shallow meta-learning Ensemble CNNs models and Deep meta-learning Ensemble CNNs models were then presented in this study. The Voting Ensemble CNNs model utilized the final prediction probabilities of the four trained CNN models. The Shallow meta-learning Ensemble CNNs models utilized the feature vectors extracted by the Deep CNNs to train classical machine learning models. The Deep meta-learning Ensemble CNNs models utilized the feature maps extracted by the Deep CNNs to train deep meta-learners (NN or FCN or Attention). The presented Deep meta-learning Ensemble CNNs models outperformed the Deep CNNs models, Conventional Ensemble CNNs models and Shallow meta-learning Ensemble CNNs models achieving state-of-the-art results on multiple evaluation metrics. This study also proposed a novel Attentive Ensemble CNNs model that learns the relative importance between the feature embeddings thereby allowing the model to focus on relevant feature embeddings while making the prediction. The improved performance of the Deep Ensemble Learning models can be attributed to the reduced bias and variance obtained as a result of increased model complexity and the synergized deep feature maps.

Future developments of the proposed approach include using 3D CT scans to train 3D CNN Ensemble models. Vision Transformers can also be experimented as base learners by replacing Deep CNNs. The proposed approach can also be used for CT scan segmentation by training on well annotated CT scans. This can enable the radiologists to locate the exact area of the infection. Activation maps of the Deep CNN base learners can be visualized to identify the part of the CT scan image that proved to be relevant for the particular classification. The proposed classification approach can be extended to more classes by increasing the variety and volume of the database.

Author Contributions: Conceptualization, J.B.T. and S.K.V.; methodology, J.B.T, S.K.V, S.M.S and A.A.J; software, J.B.T.; validation, J.B.T, S.K.V, S.M.S and A.A.J.; formal analysis, J.B.T,S.K.V, S.M.S. and A.A.J; data curation, J.B.T. and S.K.V.; writing—original draft preparation, J.B.T.;

writing—review and editing, J.B.T., S.K.V., S.M.S. and A.A.J; Funding: A.A.J; All authors have read and agreed to the published version of the manuscript

Data Availability Statement: The code of the paper is available at: <https://github.com/jibin-t-2k/Deep-Ensemble-CNNs-for-COVID-19-Chest-CT-Classification>

Conflicts of Interest: The authors declare no conflict of interest.

References

- Filchakova, O.; Dossym, D.; Ilyas, A.; Kuanysheva, T.; Abdizhamil, A.; Bukasov, R. Review of COVID-19 Testing and Diagnostic Methods. *Talanta* **2022**, *244*, 123409, doi:10.1016/j.talanta.2022.123409.
- Harmon, S.A.; Sanford, T.H.; Xu, S.; Turkbey, E.B.; Roth, H.; Xu, Z.; Stellato, E.; Plensich, G.G.; Franceschelli, G.; Girlando, C.; et al. Artificial Intelligence for the Detection of COVID-19 Pneumonia on Chest CT Using Multinational Datasets. *Nature Communications* **2020**, *11*, 4080, doi:10.1038/s41467-020-17971-2.
- Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis* **2017**, *42*, 60–88, doi:10.1016/j.media.2017.07.005.
- Alakus, T.B.; Turkoglu, I. Comparison of Deep Learning Approaches to Predict COVID-19 Infection. *Chaos, Solitons & Fractals* **2020**, *140*, 110120, doi:10.1016/j.chaos.2020.110120.
- Babukarthik, R.G.; Adiga, V.A.K.; Sambasivam, G.; Chandramohan, D.; Amudhavel, J. Prediction of COVID-19 Using Genetic Deep Learning Convolutional Neural Network (GDCNN). *IEEE Access* **2020**, *8*, 177647–177666, doi:10.1109/access.2020.3025164.
- Serte, S.; Demirel, H. Deep Learning for Diagnosis of COVID-19 Using 3D CT Scans. *Computers in Biology and Medicine* **2021**, *132*, 104306, doi:10.1016/j.compbimed.2021.104306.
- Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M.; Guo, J.; Cai, M.; Yang, J.; Li, Y.; Meng, X.; et al. A Deep Learning Algorithm Using CT Images to Screen for Corona Virus Disease (COVID-19). *European Radiology* **2021**, *31*, doi:10.1007/s00330-021-07715-1.
- Shan, F.; Gao, Y.; Wang, J.; Shi, W.; Shi, N.; Han, M.; Xue, Z.; Shen, D.; Shi, Y. Abnormal Lung Quantification in Chest CT Images of COVID-19 Patients with Deep Learning and Its Application to Severity Prediction. *Medical Physics* **2020**, *48*, doi:10.1002/mp.14609.
- Hu, S.; Gao, Y.; Niu, Z.; Jiang, Y.; Li, L.; Xiao, X.; Wang, M.; Fang, E.F.; Menpes-Smith, W.; Xia, J.; et al. Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification from CT Images. *IEEE Access* **2020**, *8*, 118869–118883, doi:10.1109/access.2020.3005510.
- Panwar, H.; Gupta, P.K.; Siddiqui, M.K.; Morales-Menendez, R.; Bhardwaj, P.; Singh, V. A Deep Learning and Grad-CAM Based Color Visualization Approach for Fast Detection of COVID-19 Cases Using Chest X-Ray and CT-Scan Images. *Chaos, Solitons & Fractals* **2020**, *140*, 110190, doi:10.1016/j.chaos.2020.110190.
- Song, Y.; Zheng, S.; Li, L.; Zhang, X.; Zhang, X.; Huang, Z.; Chen, J.; Wang, R.; Zhao, H.; Zha, Y.; et al. Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2021**, *18*, 1–1, doi:10.1109/tcbb.2021.3065361.
- Zhao, W.; Jiang, W.; Qiu, X. Deep Learning for COVID-19 Detection Based on CT Images. *Scientific Reports* **2021**, *11*, 14353, doi:10.1038/s41598-021-93832-2.
- Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A. Application of Deep Learning Technique to Manage COVID-19 in Routine Clinical Practice Using CT Images: Results of 10 Convolutional Neural Networks. *Comput. Biol. Med.* **2020**, *121*, 103795, doi:10.1016/j.compbimed.2020.103795.
- Singh, V.K.; Kolekar, M.H. Deep Learning Empowered COVID-19 Diagnosis Using Chest CT Scan Images for Collaborative Edge-

- Cloud Computing Platform. *Multimed. Tools Appl.* **2022**, *81*, 3–30. 858
15. Wang, B.; Jin, S.; Yan, Q.; Xu, H.; Luo, C.; Wei, L.; Zhao, W.; Hou, X.; Ma, W.; Xu, Z.; et al. AI-Assisted CT Imaging Analysis for COVID-19 Screening: Building and Deploying a Medical AI System. *Appl. Soft Comput.* **2021**, *98*, 106897. 859
 16. Marques, G.; Agarwal, D.; de la Torre D'v'iez, I. Automated Medical Diagnosis of COVID-19 through EfficientNet Convolutional Neural Network. *Appl. Soft Comput.* **2020**, *96*, 106691. 860
 17. Yousefzadeh, M.; Esfahanian, P.; Movahed, S.M.S.; Gorgin, S.; Rahmati, D.; Abedini, A.; Nadj, S.A.; Haseli, S.; Bakhshayesh Karam, M.; Kiani, A.; et al. Ai-Corona: Radiologist-Assistant Deep Learning Framework for COVID-19 Diagnosis in Chest CT Scans. *PLoS One* **2021**, *16*, e0250952. 861
 18. Kaya, A.; Atas, K.; Myderrizi, I. Implementation of CNN Based COVID-19 Classification Model from CT Images. In Proceedings of the 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI); 2021; pp. 201–206. 862
 19. Ibrahim, M.R.; Youssef, S.M.; Fathalla, K.M. Abnormality Detection and Intelligent Severity Assessment of Human Chest Computed Tomography Scans Using Deep Learning: A Case Study on SARS-COV-2 Assessment. *J. Ambient Intell. Humaniz. Comput.* **2021**, 1–24. 863
 20. Jin, C.; Chen, W.; Cao, Y.; Xu, Z.; Tan, Z.; Zhang, X.; Deng, L.; Zheng, C.; Zhou, J.; Shi, H.; et al. Development and Evaluation of an Artificial Intelligence System for COVID-19 Diagnosis. *Nat. Commun.* **2020**, *11*, 1–14. 864
 21. Kamel, M.A.; Abdelshafy, M.; AbdulRazek, M.; Abouelkhir, O.; Fawzy, A.; Sahlol, A.T. Efficient Classification Approach Based on Covid-19 Ct Images Analysis with Deep Features. In Proceedings of the 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS); 2021; pp. 459–464. 865
 22. Garain, A.; Basu, A.; Giampaolo, F.; Velasquez, J.D.; Sarkar, R. Detection of COVID-19 from CT Scan Images: A Spiking Neural Network-Based Approach. *Neural Comput. Appl.* **2021**, *33*, 12591–12604. 866
 23. Bolón-Canedo, V.; Alonso-Betanzos, A. Ensembles for Feature Selection: A Review and Future Trends. *Information Fusion* **2019**, *52*, 1–12, doi:10.1016/j.inffus.2018.11.008. 867
 24. Biswas, S.; Chatterjee, S.; Majee, A.; Sen, S.; Schwenker, F.; Sarkar, R. Prediction of COVID-19 from Chest CT Images Using an Ensemble of Deep Learning Models. *Applied Sciences* **2021**, *11*, 7004, doi:10.3390/app11157004. 868
 25. Das, A.K.; Ghosh, S.; Thunder, S.; Dutta, R.; Agarwal, S.; Chakrabarti, A. Automatic COVID-19 Detection from X-Ray Images Using Ensemble Learning with Convolutional Neural Network. *Pattern Analysis and Applications* **2021**, *24*, 1111–1124, doi:10.1007/s10044-021-00970-4. 869
 26. Shaik, N.S.; Cherukuri, T.K. Transfer Learning Based Novel Ensemble Classifier for COVID-19 Detection from Chest CT-Scans. *Computers in Biology and Medicine* **2022**, *141*, 105127, doi:10.1016/j.combiomed.2021.105127. 870
 27. Singh, M.; Bansal, S.; Ahuja, S.; Dubey, R.K.; Panigrahi, B.K.; Dey, N. Transfer Learning-Based Ensemble Support Vector Machine Model for Automated COVID-19 Detection Using Lung Computerized Tomography Scan Data. *Medical & Biological Engineering & Computing* **2021**, *59*, 825–839, doi:10.1007/s11517-020-02299-2. 871
 28. Kartal, E. A Comparative Analysis on Improving Covid-19 Prediction by Using Ensemble Learning Methods. In Proceedings of the Digitizing Production Systems; Springer International Publishing, 2022; pp. 3–14. 872
 29. Alrahhal, M.; K.P, S. COVID-19 Diagnostic System Using Medical Image Classification and Retrieval: A Novel Method for Image Analysis. *Comput. J.* **2021**, doi:10.1093/comjnl/bxab051. 873
 30. 13. Bougourzi, F.; Contino, R.; Distant, C.; Taleb-Ahmed, A. CNR-IEMN: A Deep Learning Based Approach to Recognise Covid-19 from CT-Scan. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2021; pp. 8568–8572. 874
 31. 14. Das, A. Adaptive UNet-Based Lung Segmentation and Ensemble Learning with CNN-Based Deep Features for Automated COVID-19 Diagnosis. *Multimed. Tools Appl.* **2022**, *81*, 5407–5441, doi:10.1007/s11042-021-11787-y. 875
 32. 15. Foysal, M.; Hossain, A.B.M.A. COVID-19 Detection from Chest CT Images Using Ensemble Deep Convolutional Neural 876

- Network. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET); 2021; pp. 1–6. 900
33. 16. Li, X.; Tan, W.; Liu, P.; Zhou, Q.; Yang, J. Classification of COVID-19 Chest CT Images Based on Ensemble Deep Learning. *J. Healthc. Eng.* **2021**, 2021, 5528441, doi:10.1155/2021/5528441. 901
902
903
34. Bahdanau, D.; Cho, K.H.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the ICLR 2015 - Conference Track Proceedings; 2015. 904
905
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems; 2017; pp. 5999–6009. 906
907
36. Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q.V.; Salakhutdinov, R. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2021; pp. 2978–2988. 908
909
910
37. Krizhevsky, A.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems; 2012; pp. 1097–1105. 911
912
38. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the Conference Track Proceedings; 2015; pp. 1–14. 913
914
39. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the Lecture Notes in Computer Science; Springer, 2014; Vol. 8689, pp. 818–833. 915
916
40. Afshar, P.; Heidarian, S.; Enshaei, N.; Naderkhani, F.; Rafiee, M.J.; Oikonomou, A.; Fard, F.B.; Samimi, K.; Plataniotis, K.N.; Mohammadi, A. COVID-CT-MD, COVID-19 Computed Tomography Scan Dataset Applicable in Machine Learning and Deep Learning. *Scientific Data* **2021**, 8, doi:10.1038/s41597-021-00900-3. 917
918
919
41. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. COVID-19 Image Data Collection: Prospective Predictions Are the Future. *arXiv:2006.11988 [cs, eess, q-bio]* **2020**. 920
921
42. Jun, M.; Cheng, G.; Yixin, W.; et al COVID-19 CT Lung and Infection Segmentation Dataset. *Zenodo* **2020**. 922
43. Morozov, S.P.; Andreychenko, A.E.; Blokhin, I.A.; Gelezhe, P.B.; Gonchar, A.P.; Nikolaev, A.E.; Pavlov, N.A.; Chernina, V.Yu.; Gomboleviskiy, V.A. MosMedData: Data Set of 1110 Chest CT Scans Performed during the COVID-19 Epidemic. *Digital Diagnostics* **2020**, 1, 49–59, doi:10.17816/dd46826. 923
924
925
44. Rahimzadeh, M.; Attar, A.; Sakhaei, S.M. A Fully Automated Deep Learning-Based Network for Detecting COVID-19 from a New and Large Lung CT Scan Dataset. *Biomedical Signal Processing and Control* **2021**, 68, 102588, doi:10.1016/j.bspc.2021.102588. 926
927
928
45. Yang, X.; He, X.; Zhao, J.; Zhang, Y.; Zhang, S.; Xie, P. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. *arXiv:2003.13865 [cs, eess, stat]* **2020**. 929
930
46. Maftouni, M.; Law, A.C.C.; Shen, B.; Kong, Z.G.; Zhou, Y.; Yazdi, N.A. A Robust Ensemble-Deep Learning Model for COVID-19 Diagnosis Based on an Integrated CT Scan Images Database. In Proceedings of the IISE Annual Conference and Expo; 2021; pp. 632–637. 931
932
933
47. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc., 2012; Vol. 25. 934
935
936
48. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* **2017**, 31, doi:10.1609/aaai.v31i1.11231. 937
938
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; pp. 770–778. 939
940

-
50. Szegedy, C.; Vincent, V.; Sergey, L.; Jonathon, S.; Zbigniew, W. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; pp. 2818–2826.
51. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; pp. 2261–2269.
52. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; pp. 1800–1807.
53. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning; 2019; pp. 10691–10700.
54. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble Deep Learning: A Review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151, doi:10.1016/j.engappai.2022.105151.
55. 19. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *Conf. Neural Inf. Process. Syst.* **2017**.
56. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the ICLR 2015 - Conference Track Proceedings.; 2015.
57. Thomas, J.B.; Devvarma, M.; Shihabudheen, K.V. Deep Ensemble Approaches for Classification of COVID-19 in Chest X-Ray Images. In Proceedings of the International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT); 2021; pp. 463–468.
58. Zhang, K.; Liu, X.; Shen, J.; Li, Z.; Sang, Y.; Wu, X.; Zha, Y.; Liang, W.; Wang, C.; Wang, K.; et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* **2020**, *181*, 1423–1433.e11, doi:10.1016/j.cell.2020.04.045.